Rethinking the stellar explosions that mark cosmic distances p. 1046

Mangrove survival in rising seas pp. 1050 & 1118

Next-generation fluorocarbons in the environment pp. 1066 & 1103

Science

$15
5 JUNE 2020
sciencemag.org

AAAS

METALLIC SOLUTION

Electrolyte-to-metal transition of electrons dissolved in ammonia pp. 1056 & 1086

CONTENTS

5 JUNE 2020 • VOLUME 368 • ISSUE 6495

PHOTO: NOAH BERGER/AP PHOTO

NEWS

IN BRIEF

1036 News at a glance

IN DEPTH

1038 Sex hormones signal why virus hits men harder
Emerging role of androgens suggests potential therapies to suppress infection *By M. Wadman*
PODCAST

1039 Blood vessel injury may spur disease's fatal second phase
Damage to vessel lining may drive mysterious clotting disorders, inflammation
By C. Matacic

1041 The pandemic's first major research scandal erupts
Critics question patient data used to challenge malaria drugs *By K. Servick and M. Enserink*

1042 Shuttered natural history museums fight for survival
But creative scientists vow recovery and move research and public programs online *By E. Pennisi*

1044 Europe bets R&D spending will bring jobs to battered economy
Horizon Europe gets €13.5 billion to spend fast, spur growth
By N. Wallace

1045 Bill would supersize NSF's budget—and role
Legislation calls for $100 billion increase, new technology directorate, and new name
By J. Mervis

FEATURES

1046 Double trouble
The brightest supernovae, used as cosmic yardsticks, have an unexpected trigger: pairs of dead stars *By D. Clery*

INSIGHTS

PERSPECTIVES

1050 Blue carbon from the past forecasts the future
Coastal mangrove ecosystems must rise above sea level to survive *By C. E. Lovelock*
REPORT p. 1118

1052 The specifics of innate immune memory
Innate immune cells develop specific memory involved in transplant rejection in mice
By J. Domínguez-Andrés and M. G. Netea
REPORT p. 1122

1053 Testicular-borne factors affect sperm fertility
Immature testicular germ cells secrete factors that influence sperm maturation in mice
By T. Lord and J. M. Oatley
REPORT p. 1132

1055 Deep-ocean seafloor islands of plastics
The processes controlling sediment transport also concentrate microplastics *By D. Mohrig*
REPORT p. 1140

1056 The link between electrolytes and metals
Photoelectron spectroscopy maps a gradual transition for solvated electrons in ammonia *By C. M. Isborn*
RESEARCH ARTICLE p. 1086

1057 Unblinding with infrared nanosensors
Gene therapy and nanotechnology come together to fight degenerative blindness
By K. Franke and A. Vlasits
REPORT p. 1108

1058 Follow the barcoded microbes
Genetically engineered spores give object provenance technology new avenues
By J. Nivala
REPORT p. 1135

1060 Serology assays to manage COVID-19
Measurement of antibodies to SARS-CoV-2 will improve disease management if used correctly
By F. Krammer and V. Simon

1062 Donald Kennedy (1931–2020)
Eminent public intellectual who served science and society *By G. C. Daily and P. R. Ehrlich*

POLICY FORUM

1063 Which interventions work best in a pandemic?
We can exploit randomized controlled trials, compartmental models, and spillovers
By J. Haushofer and C. J. E. Metcalf

Published by AAAS

1066 Filling gaps in science exposes gaps in chemical regulation
Examination of U.S. and EU regulatory systems raises more questions than answers
By S. C. Gold and W. E. Wagner
REPORT p. 1103

BOOKS *ET AL.*

1071 Colonialism and its consequences
A ravaged Arctic ecosystem serves as a warning of the perils of human advancement
By M. E. Hannibal

1072 Where logic meets emotion
A computer scientist gets candid about her quest to bring empathy to artificial intelligence *By V. Venkatraman*

LETTERS

1074 Recovery hopes for the world's rarest primate
By H. Liu et al.

1074 Cetaceans under threat in South China Sea
By S. Li et al.

1075 Tenure and promotion after the pandemic
By M. Htun

1075 Technical Comment abstracts

RESEARCH

IN BRIEF

1076 From *Science* and other journals

REVIEW

1079 Computer science
There's plenty of room at the Top: What will drive computer performance after Moore's law? *C. E. Leiserson* et al.
REVIEW SUMMARY; FOR FULL TEXT: DX.DOI.ORG/10.1126/SCIENCE.AAM9744

RESEARCH ARTICLES

1080 Plasmid evolution
Unexpected conservation and global transmission of agrobacterial virulence plasmids *A. J. Weisberg* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DX.DOI.ORG/10.1126/SCIENCE.ABA5256

1081 Structural biology
The structure of human CST reveals a decameric assembly bound to telomeric DNA *C. J. Lim* et al.

1086 Chemical physics
Photoelectron spectra of alkali metal–ammonia microjets: From blue electrolyte to bronze metal *T. Buttersack* et al.
PERSPECTIVE p. 1056

1091 Wearable devices
Giant thermopower of ionic gelatin near room temperature *C.-G. Han* et al.

REPORTS

1098 Organic chemistry
Iridium-catalyzed acid-assisted asymmetric hydrogenation of oximes to hydroxylamines *J. Mas-Roselló* et al.

1103 Analytical chemistry
Nontargeted mass-spectral detection of chloroperfluoropolyether carboxylates in New Jersey soils *J. W. Washington* et al.
POLICY FORUM p. 1066

1108 Neuroscience
Restoring light sensitivity using tunable near-infrared sensors *D. Nelidova* et al.
PERSPECTIVE p. 1057

1114 Topological optics
A fractional corner anomaly reveals higher-order topology *C. W. Peterson* et al.

1118 Paleoecology
Thresholds of mangrove survival under rapid sea level rise *N. Saintilan* et al.
PERSPECTIVE p. 1050

1122 Immunology
PIRs mediate innate myeloid cell memory to nonself MHC molecules
H. Dai et al.
PERSPECTIVE p. 1052

1127 Cancer
MTOR signaling orchestrates stress-induced mutagenesis, facilitating adaptive evolution in cancer *A. Cipponi* et al.

1132 Reproductive biology
NELL2-mediated lumicrine signaling through OVCH2 is required for male fertility
D. Kiyozumi et al.
PERSPECTIVE p. 1053

1135 Molecular biology
Barcoded microbial system for high-resolution object provenance
J. Qian et al.
PERSPECTIVE p. 1058; PODCAST

1140 Plastic pollution
Seafloor microplastic hotspots controlled by deep-sea circulation
I. A. Kane et al.
PERSPECTIVE p. 1055

DEPARTMENTS

1035 Editorial
Patents, economics, and pandemics
By Will Zerhouni et al.

1150 Working Life
No place like home *By Guilherme Santos*

ON THE COVER

Artistic abstraction of how ammonia solutions change with increasing excess electron concentration. Microjets of liquid ammonia can be used to track how an electrically insulating liquid becomes a metal, with an accompanying color change from transparent blue through deeper blue to bronze. Mapping energy levels spectroscopically and theoretically reveals a gradual transition from electrolyte to liquid metal. See pages 1056 and 1086.
*Illustration: C. Bickel/*Science

Science Staff 1034
New Products 1146
Science Careers 1147

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Copyright © 2020 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $2148; Foreign postage extra: Air assist delivery: $98. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**
Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling available from backissues.sciencemag.org; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

PHOTO: 1441147448/SHUTTERSTOCK

Published by AAAS

Science AAAS

Editor-in-Chief Holden Thorp, hthorp@aaas.org

Executive Editor Monica M. Bradford

Editors, Research Valda Vinson, Jake S. Yeston **Editor, Insights** Lisa D. Chong

DEPUTY EDITORS Julia Fahrenkamp-Uppenbrink (UK), Stella M. Hurtley (UK), Phillip D. Szuromi, Sacha Vignieri **SR. EDITORIAL FELLOW** Andrew M. Sugden (UK) **SR. EDITORS** Gemma Alderton (UK), Caroline Ash (UK), Brent Grocholski, Pamela J. Hines, Marc S. Lavine (Canada), Ian S. Osborne (UK), Beverly A. Purnell, L. Bryan Ray, H. Jesse Smith, Keith T. Smith (UK), Jelena Stajic, Peter Stern (UK), Valerie B. Thompson, Brad Wible, Laura M. Zahn **ASSOCIATE EDITORS** Michael A. Funk, Priscilla N. Kelly, Tage S. Rai, Seth Thomas Scanlon (UK), Yury V. Suleymanov **LETTERS EDITOR** Jennifer Sills **LEAD CONTENT PRODUCTION EDITORS** Harry Jach, Lauren Kmec **CONTENT PRODUCTION EDITORS** Amelia Beyna, Jeffrey E. Cook, Chris Filiatreau, Julia Katris, Nida Masiulis, Suzanne M. White **SR. EDITORIAL COORDINATORS** Carolyn Kyle, Beverly Shields **EDITORIAL COORDINATORS** Aneera Dobbins, Joi S. Granger, Jeffrey Hearn, Lisa Johnson, Maryrose Madrid, Ope Martins, Shannon McMahon, Jerry Richardson, Hilary Stewart (UK), Alana Warnke, Alice Whaley (UK), Anita Wynn **PUBLICATIONS ASSISTANTS** Jeremy Dow, Alexander Kief, Ronmel Navas, Brian White **EXECUTIVE ASSISTANT** Jessica Slater **ASI DIRECTOR, OPERATIONS** Janet Clements (UK) **ASI SR. OFFICE ADMINISTRATOR** Jessica Waldock (UK)

News Editor Tim Appenzeller

NEWS MANAGING EDITOR John Travis **INTERNATIONAL EDITOR** Martin Enserink **DEPUTY NEWS EDITORS** Elizabeth Culotta, Lila Guterman, David Grimm, Eric Hand (Europe), David Malakoff **SR. CORRESPONDENTS** Daniel Clery (UK), Jon Cohen, Jeffrey Mervis, Elizabeth Pennisi **ASSOCIATE EDITORS** Jeffrey Brainard, Catherine Matacic **NEWS REPORTERS** Adrian Cho, Jennifer Couzin-Frankel, Jocelyn Kaiser, Kelly Servick, Robert F. Service, Erik Stokstad, Paul Voosen, Meredith Wadman **INTERNS** Eva Frederick, Rodrigo Perez Ortega **CONTRIBUTING CORRESPONDENTS** Warren Cornwall, Ann Gibbons, Mara Hvistendahl, Sam Kean, Eli Kintisch, Kai Kupferschmidt (Berlin), Andrew Lawler, Mitch Leslie, Eliot Marshall, Virginia Morell, Dennis Normile (Shanghai), Elisabeth Pain (Careers), Charles Piller, Michael Price, Tania Rabesandratana (Barcelona), Emily Underwood, Gretchen Vogel (Berlin), Lizzie Wade (Mexico City) **CAREERS** Donisha Adams, Rachel Bernstein (Editor), Katie Langin (Acting Editor) **COPY EDITORS** Julia Cole (Senior Copy Editor), Cyra Master (Copy Chief) **ADMINISTRATIVE SUPPORT** Meagan Weiland

Creative Director Beth Rakouskas

DESIGN MANAGING EDITOR Marcy Atarod **GRAPHICS MANAGING EDITOR** Alberto Cuadra **PHOTOGRAPHY MANAGING EDITOR** William Douthitt **WEB CONTENT STRATEGY MANAGER** Kara Estelle-Powers **DESIGN EDITOR** Chrystal Smith **DESIGNER** Christina Aycock **GRAPHICS EDITOR** Nirja Desai **INTERACTIVE GRAPHICS EDITOR** Xing Liu **SENIOR SCIENTIFIC ILLUSTRATORS** Valerie Altounian, Chris Bickel **SCIENTIFIC ILLUSTRATOR** Alice Kitterman **SENIOR GRAPHICS SPECIALISTS** Holly Bishop, Nathalie Cary **SENIOR PHOTO EDITOR** Emily Petersen **PHOTO EDITOR** Kaitlyn Dolan

Chief Executive Officer and Executive Publisher Sudip Parikh

Publisher, Science Family of Journals Bill Moran

DIRECTOR, BUSINESS SYSTEMS AND FINANCIAL ANALYSIS Randy Yi **DIRECTOR, BUSINESS OPERATIONS & ANALYSIS** Eric Knott **DIRECTOR OF ANALYTICS** Enrique Gonzales **MANAGER, BUSINESS OPERATIONS** Jessica Tierney **SENIOR BUSINESS ANALYST** Cory Lipman, Meron Kebede **FINANCIAL ANALYST** Alexander Lee **ADVERTISING SYSTEM ADMINISTRATOR** Tina Burks **SENIOR SALES COORDINATOR** Shirley Young **DIGITAL/PRINT STRATEGY MANAGER** Jason Hillman **QUALITY TECHNICAL MANAGER** Marcus Spiegler **ASSISTANT MANAGER DIGITAL/PRINT** Rebecca Doshi **SENIOR CONTENT SPECIALISTS** Steve Forrester, Jacob Hedrick, Antoinette Hodal, Lori Murphy **DIGITAL PRODUCTION MANAGER** Lisa Stanford **CONTENT SPECIALIST** Kimberley Oster **ADVERTISING PRODUCTION OPERATIONS MANAGER** Deborah Tompkins **DESIGNER, CUSTOM PUBLISHING** Jeremy Huntsinger **SR. TRAFFIC ASSOCIATE** Christine Hall **SPECIAL PROJECTS ASSOCIATE** Sarah Dhere

ASSOCIATE DIRECTOR, BUSINESS DEVELOPMENT Justin Sawyers **GLOBAL MARKETING MANAGER** Allison Pritchard **DIGITAL MARKETING MANAGER** Aimee Aponte **JOURNALS MARKETING MANAGER** Shawana Arnold **MARKETING ASSOCIATES** Tori Velasquez, Mike Romano, Ashley Hylton **DIGITAL MARKETING SPECIALIST** Asleigh Rojanavongse **SENIOR DESIGNER** Kim Huynh

DIRECTOR AND SENIOR EDITOR, CUSTOM PUBLISHING Sean Sanders **ASSISTANT EDITOR, CUSTOM PUBLISHING** Jackie Oberst

DIRECTOR, PRODUCT & PUBLISHING DEVELOPMENT Chris Reid **DIRECTOR, BUSINESS STRATEGY AND PORTFOLIO MANAGEMENT** Sarah Whalen **ASSOCIATE DIRECTOR, PRODUCT MANAGMENT** Kris Bishop **SR. PRODUCT ASSOCIATE** Robert Koepke **DIGITAL PRODUCT STRATEGIST** Michael Hardesty **SPJ ASSOCIATE** Samantha Bruno Fuller

DIRECTOR, INSTITUTIONAL LICENSING Iquo Edim **ASSOCIATE DIRECTOR, RESEARCH & DEVELOPMENT** Elisabeth Leonard **MARKETING MANAGER** Kess Knight **SENIOR INSTITUTIONAL LICENSING MANAGER** Ryan Rexroth **INSTITUTIONAL LICENSING MANAGER** Marco Castellan **MANAGER, AGENT RELATIONS & CUSTOMER SUCCESS** Judy Lillibridge **SENIOR OPERATIONS ANALYST** Lana Guz **FULFILLMENT COORDINATOR** Melody Stringer **SALES COORDINATOR** Josh Haverlock

DIRECTOR, GLOBAL SALES Tracy Holmes **US EAST COAST AND MID WEST SALES** Stephanie O'Connor **US WEST COAST SALES** Lynne Stickrod **US SALES MANAGER, SCIENCE CAREERS** Claudia Paulsen-Young **US SALES REP, SCIENCE CAREERS** Tracy Anderson **ASSOCIATE DIRECTOR, ROW** Roger Goncalves **SALES REP, ROW** Sarah Lelarge **SALES ADMIN ASSISTANT, ROW** Bryony Cousins **DIRECTOR OF GLOBAL COLLABORATION AND ACADEMIC PUBLISHING RELATIONS, ASIA** Xiaoying Chu **ASSOCIATE DIRECTOR, INTERNATIONAL COLLABORATION** Grace Yao **SALES MANAGER** Danny Zhao **MARKETING MANAGER** Kilo Lan **ASCA CORPORATION, JAPAN** Kaoru Sasaki (Tokyo), Miyuki Tani (Osaka) **COLLABORATION/CUSTOM PUBLICATIONS/JAPAN** Adarsh Sandhu

DIRECTOR, COPYRIGHT, LICENSING AND SPECIAL PROJECTS Emilie David **RIGHTS AND LICENSING COORDINATOR** Jessica Adams **RIGHTS AND PERMISSIONS ASSOCIATE** Elizabeth Sandler **CONTRACTS AND LICENSING ASSOCIATE** Lili Catlett

---

MAIN HEADQUARTERS
Science/AAAS
1200 New York Ave. NW
Washington, DC 20005

SCIENCE INTERNATIONAL
Clarendon House
Clarendon Road
Cambridge, CB2 8FH, UK

SCIENCE CHINA
Room 1004, Culture Square
No. 59 Zhongguancun St.
Haidian District, Beijing, 100872

SCIENCE JAPAN
ASCA Corporation
Sibaura TY Bldg. 4F, 1-14-5
Shibaura Minato-ku
Tokyo, 108-0073 Japan

EDITORIAL
science_editors@aaas.org

NEWS
science_news@aaas.org

INFORMATION FOR AUTHORS
sciencemag.org/authors/science-information-authors

REPRINTS AND PERMISSIONS
sciencemag.org/help/reprints-and-permissions

MEDIA CONTACTS
scipak@aaas.org

MULTIMEDIA CONTACTS
SciencePodcast@aaas.org
ScienceVideo@aaas.org

INSTITUTIONAL SALES AND SITE LICENSES
sciencemag.org/librarian

PRODUCT ADVERTISING & CUSTOM PUBLISHING
advertising.sciencemag.org/products-services
science_advertising@aaas.org

CLASSIFIED ADVERTISING
advertising.sciencemag.org/science-careers
advertise@sciencecareers.org

JOB POSTING CUSTOMER SERVICE
employers.sciencecareers.org
support@sciencecareers.org

MEMBERSHIP AND INDIVIDUAL SUBSCRIPTIONS
sciencemag.org/subscriptions

MEMBER BENEFITS
aaas.org/membercentral

AAAS BOARD OF DIRECTORS
CHAIR Steven Chu
PRESIDENT Claire M. Fraser
PRESIDENT-ELECT Susan G. Amara
TREASURER Carolyn N. Ainslie
CHIEF EXECUTIVE OFFICER Sudip Parikh
BOARD Cynthia M. Beall, Rosina M. Bierbaum, Ann Bostrom, Stephen P.A. Fodor, S. James Gates, Jr., Laura H. Greene, Kaye Husbands Fealing, Maria M. Klawe, Robert B. Millard, Alondra Nelson, William D. Provine

BOARD OF REVIEWING EDITORS (Statistics board members indicated with **(S)**)

Adriano Aguzzi, *U. Hospital Zürich*
Takuzo Aida, *U. of Tokyo*
Leslie Aiello, *Wenner-Gren Foundation*
Judith Allen, *U. of Manchester*
Sebastian Amigorena, *Institut Curie*
James Analytis, *U. of California, Berkeley*
Paola Arlotta, *Harvard U.*
Johan Auwerx, *EPFL*
David Awschalom, *U. of Chicago*
Clare Baker, *U. of Cambridge*
Nenad Ban, *ETH Zürich*
Franz Bauer, *Pontificia Universidad Católica de Chile*
Ray H. Baughman, *U. of Texas at Dallas*
Peter Bearman, *Columbia U.*
Carlo Beenakker, *Leiden U.*
Yasmine Belkaid, *NIAID, NIH*
Philip Benfey, *Duke U.*
Gabriele Bergers, *VIB*
Bradley Bernstein, *Mass. General Hospital*
Alessandra Biffi, *Harvard Med. School*
Peer Bork, *EMBL*
Chris Bowler, *École Normale Supérieure*
Ian Boyd, *U. of St. Andrews*
Emily Brodsky, *U. of California, Santa Cruz*
Ron Brookmeyer, *U. of California, Los Angeles* **(S)**
Christian Büchel, *UKE Hamburg*
Dennis Burton, *Scripps Research*
Carter Tribley Butts, *U. of California, Irvine*
György Buzsáki, *New York U. School of Med.*
Blanche Capel, *Duke U.*
Annmarie Carlton, *U. of California, Irvine*
Nick Chater, *U. of Warwick*
Zhijian Chen, *UT Southwestern Med. Ctr.*
Ib Chorkendorff, *Denmark TU*
James J. Collins, *MIT*
Robert Cook-Deegan, *Arizona State U.*
Alan Cowman, *Walter & Eliza Hall Inst.*
Carolyn Coyne, *U. of Pittsburgh*
Roberta Croce, *VU Amsterdam*
Jeff L. Dangl, *U. of North Carolina*
Tom Daniel, *U. of Washington*
Chiara Daraio, *Caltech*
Nicolas Dauphas, *U. of Chicago*
Frans de Waal, *Emory U.*
Claude Desplan, *New York U.*
Sandra Díaz, *Universidad Nacional de Córdoba*
Ulrike Diebold, *TU Wien*
Hong Ding, *Inst. of Physics, CAS*
Jennifer Dionne, *Stanford U.*
Dennis Discher, *U. of Penn.*
Gerald Dorn, *Washington U. in St. Louis*
Jennifer A. Doudna, *U. of California, Berkeley*
Bruce Dunn, *U. of California, Los Angeles*
William Dunphy, *Caltech*
Christopher Dye, *U. of Oxford*
Todd Ehlers, *U. of Tübingen*
Jennifer Elisseeff, *Johns Hopkins U.*
Tim Elston, *U. of North Carolina*
Andrea Encalada, *U. San Francisco de Quito*
Nader Engheta, *U. of Penn.*
Karen Ersche, *U. of Cambridge*
Barry Everitt, *U. of Cambridge*
Vanessa Ezenwa, *U. of Georgia*
Michael Feuer, *The George Washington U.*
Toren Finkel, *U. of Pittsburgh Med. Ctr.*
Gwenn Flowers, *Simon Fraser U.*
Peter Fratzl, *Max Planck Inst. Potsdam*
Elaine Fuchs, *Rockefeller U.*
Eileen Furlong, *EMBL*
Jay Gallagher, *U. of Wisconsin*
Daniel Geschwind, *U. of California, Los Angeles*
Karl-Heinz Glassmeier, *TU Braunschweig*
Ramon Gonzalez, *U. of South Florida*
Elizabeth Grove, *U. of Chicago*
Nicolas Gruber, *ETH Zürich*
Hua Guo, *U. of New Mexico*
Kip Guy, *U. of Kentucky College of Pharmacy*
Taekjip Ha, *Johns Hopkins U.*
Christian Haass, *Ludwig Maximilians U.*
Sharon Hammes-Schiffer, *Yale U.*
Wolf-Dietrich Hardt, *ETH Zürich*
Louise Harra, *U. College London*
Jian He, *Clemson U.*
Carl-Philipp Heisenberg, *IST Austria*
Ykä Helariutta, *U. of Cambridge*
Janet G. Hering, *Eawag*
Hans Hilgenkamp, *U. of Twente*
Kai-Uwe Hinrichs, *U. of Bremen*
Lora Hooper, *UT Southwestern Med. Ctr.*
Fred Hughson, *Princeton U.*
Randall Hulet, *Rice U.*
Auke Ijspeert, *EPFL*
Akiko Iwasaki, *Yale U.*
Stephen Jackson, *USGS and U. of Arizona*
Kai Johnsson, *EPFL*
Peter Jonas, *IST Austria*
Matt Kaeberlein, *U. of Washington*
William Kaelin Jr., *Dana-Farber Cancer Inst.*
Daniel Kammen, *U. of California, Berkeley*
V. Narry Kim, *Seoul Nat. U.*
Robert Kingston, *Harvard Med. School*
Nancy Knowlton, *Smithsonian Institution*
Etienne Koechlin, *École Normale Supérieure*
Alex L. Kolodkin, *Johns Hopkins U.*
Julija Krupic, *U. of Cambridge*
Thomas Langer, *Max Planck Inst. Cologne*
Mitchell A. Lazar, *U. of Penn.*
Ottoline Leyser, *U. of Cambridge*
Wendell Lim, *U. of California, San Francisco*
Jianguo Liu, *Michigan State U.*
Luis Liz-Marzán, *CIC biomaGUNE*
Jonathan Losos, *Washington U. in St. Louis*
Ke Lu, *Chinese Acad. of Sciences*
Christian Lüscher, *U. of Geneva*
Jean Lynch-Stieglitz, *Georgia Inst. of Tech.*
Fabienne Mackay, *QIMR Berghofer*
Anne Magurran, *U. of St. Andrews*
Oscar Marín, *King's College London*
Charles Marshall, *U. of California, Berkeley*
Christopher Marx, *U. of Idaho*
Geraldine Masson, *CNRS*
C. Robertson McClung, *Dartmouth College*
Rodrigo Medellín, *U. Nacional Autónoma de México*
Graham Medley, *London School of Hygiene & Tropical Med.*
Jane Memmott, *U. of Bristol*
Baoxia Mi, *U. of California, Berkeley*
Edward Miguel, *U. of California, Berkeley*
Tom Misteli, *NCI, NIH*
Yasushi Miyashita, *U. of Tokyo*
Alison Motsinger-Reif, *NIEHS, NIH* **(S)**
Daniel Nettle, *Newcastle U.*
Daniel Neumark, *U. of California, Berkeley*
Beatriz Noheda, *U. of Groningen*
Helga Nowotny, *Vienna Science, Research & Tech. Fund*
Rachel O'Reilly, *U. of Birmingham*
Harry Orr, *U. of Minnesota*
Pilar Ossorio, *U. of Wisconsin*
Andrew Oswald, *U. of Warwick*
Isabella Pagano, *Istituto Nazionale di Astrofisica*
Margaret Palmer, *U. of Maryland*
Elizabeth Levy Paluck, *Princeton U.*
Jane Parker, *Max Planck Inst. Cologne*
Giovanni Parmigiani, *Dana-Farber Cancer Inst.* **(S)**
Samuel Pfaff, *Salk Inst. for Biological Studies*
Julie Pfeiffer, *UT Southwestern Med. Ctr.*
Matthieu Piel, *Institut Curie*
Kathrin Plath, *U. of California, Los Angeles*
Martin Plenio, *Ulm U.*
Katherine Pollard, *U. of California, San Francisco*
Elvira Poloczanska, *Alfred-Wegener-Inst.*
Julia Pongratz, *Ludwig Maximilians U.*
Philippe Poulin, *CNRS*
Jonathan Pritchard, *Stanford U.*
Félix A. Rey, *Institut Pasteur*
Trevor Robbins, *U. of Cambridge*
Joeri Rogelj, *Imperial College London*
Amy Rosenzweig, *Northwestern U.*
Mike Ryan, *U. of Texas at Austin*
Mitinori Saitou, *Kyoto U.*
Shimon Sakaguchi, *Osaka U.*
Miquel Salmeron, *Lawrence Berkeley Nat. Lab*
Nitin Samarth, *Penn. State U.*
Jürgen Sandkühler, *Med. U. of Vienna*
Alexander Schier, *Harvard U.*
Wolfram Schlenker, *Columbia U.*
Susannah Scott, *U. of California, Santa Barbara*
Rebecca Sear, *London School of Hygiene & Tropical Med.*
Vladimir Shalaev, *Purdue U.*
Jie Shan, *Cornell U.*
Beth Shapiro, *U. of California, Santa Cruz*
Jay Shendure, *U. of Washington*
Steve Sherwood, *U. of New South Wales*
Brian Shoichet, *U. of California, San Francisco*
Robert Siliciano, *Johns Hopkins U. School of Med.*
Lucia Sivilotti, *U. College London*
Alison Smith, *John Innes Centre*
Richard Smith, *U. of North Carolina* **(S)**
Mark Smyth, *QIMR Berghofer*
Pam Soltis, *U. of Florida*
John Speakman, *U. of Aberdeen*
Tara Spires-Jones, *U. of Edinburgh*
Allan C. Spradling, *Carnegie Institution for Science*
V. S. Subrahmanian, *Dartmouth College*
Ira Tabas, *Columbia U.*
Sarah Teichmann, *Wellcome Sanger Inst.*
Rocio Titiunik, *Princeton U.*
Shubha Tole, *Tata Inst. of Fundamental Research*
Wim van der Putten, *Netherlands Inst. of Ecology*
Reinhilde Veugelers, *KU Leuven*
Bert Vogelstein, *Johns Hopkins U.*
Kathleen Vohs, *U. of Minnesota*
David Wallach, *Weizmann Inst. of Science*
Jane-Ling Wang, *U. of California, Davis* **(S)**
David Waxman, *Fudan U.*
Jonathan Weissman, *U. of California, San Francisco*
Chris Wikle, *U. of Missouri* **(S)**
Terrie Williams, *U. of California, Santa Cruz*
Ian A. Wilson, *Scripps Research* **(S)**
Yu Xie, *Princeton U.*
Jan Zaanen, *Leiden U.*
Kenneth Zaret, *U. of Penn. School of Med.*
Jonathan Zehr, *U. of California, Santa Cruz*
Xiaowei Zhuang, *Harvard U.*
Maria Zuber, *MIT*

Science serves as a forum for discussion of important issues related to the advancement of science by publishing material on which a consensus has been reached as well as including the presentation of minority or conflicting points of view. Accordingly, all articles published in *Science*—including editorials, news and comment, and book reviews—are signed and reflect the individual views of the authors and not official points of view adopted by AAAS or the institutions with which the authors are affiliated.

Published by AAAS

EDITORIAL

Patents, economics, and pandemics

As coronavirus disease 2019 (COVID-19) has quickly killed hundreds of thousands across an unprepared world, destroyed the livelihoods of millions, and cost trillions of dollars, nations must now expand their mindsets and focus not only on overcoming the next phases but also on long-range strategies—for this and for future pandemics. The world needs a permanent preparedness enterprise to engage in a sustained effort to avert pandemics, and provide for the affordable and widespread administration of vaccines and therapies when they are discovered.

Governments, academia, and industry tend to forget the threat of infectious diseases between pandemics, but when they emerge, multibillion-dollar crash programs are quickly put into place. This boom-and-bust cycle in infectious diseases research and development (R&D) has limited attractiveness to scientists, thus preventing real progress. COVID-19 has mobilized the scientific community to generate hundreds of drugs, diagnostics, and vaccines in various stages of development, but this enthusiasm will quickly evaporate once the crisis is over and funding dries up. Existing governmental and market mechanisms fail to protect society against present and future public health threats, leading to chronic underinvestment in infectious diseases in the private sector and causing many companies to quietly abandon the field. Nonprofit organizations try to fill the gap but are unlikely to meet public health needs. Without sustained government effort and investment, the world will be unprepared for pandemics.

Another reason for the market failure in pandemic protection is an inherent misalignment between economic incentives driving industry versus benefits to the public. Private innovation is driven by the patent system. Pharmaceutical companies, responsive to their shareholders, typically invest in therapies for conditions that will predictably maximize profits during the life of a patent. Patents give a time-limited exclusivity to the innovator who can then set premium pricing that maximizes the return on R&D investment. Such pricing can hinder wide dissemination once vaccines or therapies are developed, often leaving many patients unable to afford these products. Tension therefore exists between the need to generate affordable products that preserve human health and the need of innovators to be appropriately rewarded for their risk taking. This is more critical for vaccines as the individual vaccinee is not the only beneficiary. Indirect positive externalities accrue to the whole of society through faster herd immunity and more rapid economic recovery. In economic terms, this means that such products should be priced as close as possible to marginal costs, a proposition normally unattractive to drug innovators.

How is it possible to resolve this dilemma? A system of ex ante economic rewards should be created based on specific innovation and product development milestones. Governments should contribute not only emergency grant awards (which suffer from the need to pick winners and losers prospectively with scant evidence of effectiveness) but also sustained, predictable, and prospective achievement rewards or even prizes based on the value of explicitly achieved preclinical, clinical, manufacturing, and distribution milestones, including advanced market commitments when necessary to lessen the risk for the innovators who deliver. In exchange, prior to public funds being granted—and to avoid conflicts later—companies that accept taxpayer support would agree to make pandemic countermeasures available to the public rapidly, including widespread production and dissemination with reasonable margins. Today, industry practices such a system of milestone-based payments with their smaller R&D partners. Why should governments not do the same with industry for pandemics?

> **"...economic rewards should be created based on specific innovation... milestones."**

Similar proposals have been advanced in Europe, with the potential to license discoveries internationally. COVID-19 affects all of humanity, and the strategy should not be geographically restricted but global. Why not join forces? At the upcoming G7 and G20 meetings, governments should create and fund an international organization akin to a global institute of health to manage staged funding through calls for proposals from all sectors. This would sustain the R&D needed to establish rapidly scalable platforms and supply chains of diagnostics, drugs, antibodies, and vaccines against current and future pandemic threats. In exchange, innovators would be fairly rewarded, and their discoveries made available at affordable prices to people around the world. Creating such a sustainable R&D enterprise would be something positive to emerge from the wreckage of COVID-19.

–Will Zerhouni, Gary J. Nabel*, Elias Zerhouni†

Will Zerhouni is an intellectual property attorney, economist, entrepreneur, and former Assistant United States Attorney. willzerhouni@gmail.com

Gary J. Nabel is the chief scientific officer of Sanofi Global R&D, Cambridge, MA, USA, and is the former and founding director of the Vaccine Research Center at the U.S. National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA. gary.nabel@sanofi.com

Elias Zerhouni is a professor at Johns Hopkins University, Baltimore, MD, USA, and is the former director of the National Institutes of Health and former President of Global R&D for Sanofi Global R&D, Cambridge, MA, USA. eliaszerhouni1@gmail.com

*Employee and shareholder of Sanofi (active in COVID-19 R&D). The views expressed here reflect my personal opinion and are independent of policies/positions of this organization. †Board member and shareholder of Danaher and shareholder of Sanofi (both involved in COVID-19 R&D).

10.1126/science.abc7472

Published by AAAS

NEWS

> “We need to get it down to simmer before we take the lid off, and it’s too early.”
>
> **Respiratory scientist Calum Semple**, to BBC, joining others on a scientific advisory panel who voiced worries about easing U.K. lockdowns for COVID-19 this week before testing and case tracing are fully in place.

As protesters demonstrated on 1 June in New York City’s Times Square and other U.S. cities against the killing of an unarmed black man by Minneapolis police, public health specialists worried that the large crowds, in which many people did not wear masks, would quicken the spread of COVID-19.

IN BRIEF

Edited by **Jeffrey Brainard**

DISPATCHES FROM THE PANDEMIC

Yemen faces COVID-19 disaster

HUMANITARIAN AID | Yemen, already reeling from 5 years of civil war and a massive cholera outbreak, faces a catastrophic prognosis for deaths from COVID-19 unless international donors contribute to a depleted assistance fund for public health, the United Nations says. Based on modeling by a group at Imperial College London, the World Health Organization is bracing for the novel coronavirus to infect about half of Yemen’s population of about 30 million and kill an estimated 30,000 to 40,000. Official case numbers remain low: As of 1 June, the country had reported 358 cases and 85 deaths. But there are unofficial reports of hundreds more cases; rival health ministries in Aden and Sana’a have accused each other of intentionally underreporting the extent of COVID-19 in the areas they control. Even after raising $1.35 billion from donors on 2 June, the United Nations is still short nearly $1 billion in funding needed to sustain humanitarian programs in Yemen for the rest of 2020.

BY THE NUMBERS

105,644

U.S. deaths from COVID-19, as of 2 June. The total, the world’s highest, surpassed 100,000 on 28 May. Worldwide, deaths have reached more than 377,000 (Johns Hopkins University).

Trump confirms WHO exit

POLICY | President Donald Trump on 29 May said he would follow through on his threat to withdraw the United States from the World Health Organization (WHO), which he says has mishandled the coronavirus pandemic. He vowed to redistribute some of its allocated funding directly to groups it supports. Public health advocates blasted the decision, saying it could hamper a broad range of disease-fighting efforts, and noting it came just 11 days after Trump gave WHO 30 days to respond to a U.S. request for reforms. Legal experts say the United States is still obligated to pay WHO some $60 million in 2020 membership dues, and they note that an exit cannot be finalized for at least 1 year. If Trump loses the November election, the next president could reverse the decision.

PHOTO: SCOTT HEINS/GETTY IMAGES

Published by AAAS

China CDC head wants changes

PUBLIC HEALTH | The head of China's national public health agency last week asked the government to give it more decision-making power and reduce "micromanagement" by health administrators to address shortcomings revealed by the COVID-19 pandemic. Gao Fu (also known as George Gao), director of the Chinese Center for Disease Control and Prevention, offered a pointed critique of existing bureaucracies in an interview with *China Youth Daily*, an official government newspaper. During the crisis, the agency's experts could make recommendations but were not included in key decisions about the response, Gao said. Responsibility for public health is splintered among government agencies that don't coordinate their work, he added.

The other guy's COVID-19

OPINION RESEARCH | Americans think they are less likely than peers to contract COVID-19, even though many worry about the danger it poses to the broader population. In surveys in March and April, psychologists at University College London found that participants rated the pandemic as a large danger to others (74 on a scale of 100, on average, with 100 representing "extreme danger") but not to themselves (2.8 on a five-point scale, with five representing "much less likely" to get sick). Despite this apparent paradox, most respondents followed government advice on social distancing and other precautions, according to a preprint posted on the PsyArXiv server on 29 May. Those who were more worried about others were more likely to follow public health advice, the survey found, even though "optimism bias" about one's own prospects often leads people to take undue risks.

Panel faults U.K. test statistics

EPIDEMIOLOGY | The head of a U.K. agency that oversees government statistics this week criticized Secretary of State for Health and Social Care Matt Hancock for releasing numbers on the volume of COVID-19 testing that were inflated "at the expense of understanding." The figures mix completed lab results with uncompleted tests in which nose swabs were mailed for analysis, wrote David Norgrove, chairman of the UK Statistics Authority, in a 2 June letter to Hancock. The practice yields an artificially low rate of positive results, Norgrove said.

SCIENCEMAG.ORG/TAGS/CORONAVIRUS
Read additional *Science* coverage of the pandemic.

INFECTIOUS DISEASES

Congo tallies new Ebola cases

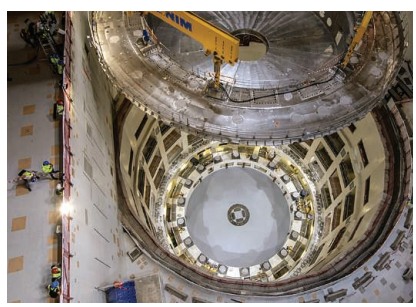
An apparently new Ebola outbreak has surfaced in the Democratic Republic of the Congo (DRC), just as the country was close to declaring the disease under control. On 1 June, the DRC's Ministry of Health said eight people in Mbandaka, a city in Équateur province, had hemorrhagic fever and four had died. Blood samples in three of the cases tested positive for the virus. In 2018, the same region quelled an Ebola outbreak that lasted 3 months. No evidence suggests this new outbreak is linked to another, ongoing one in the DRC's North Kivu province, 1000 kilometers away, that began just as the 2018 Équateur outbreak ended. North Kivu has not recorded a new case for 3 weeks. Ebola virus regularly jumps from animals to humans, and this is the DRC's 11th recorded outbreak.

Iranian nuclear work scuttled

NONPROLIFERATION | Ratcheting up its pressure campaign against Iran, the United States announced last week it will cancel waivers of economic and legal sanctions that had allowed foreign organizations to help Iran reconfigure a nuclear reactor to eliminate a proliferation risk. Originally designed for research and to produce isotopes, Iran's Arak heavy water reactor could accumulate enough plutonium in spent fuel to produce one or two bombs a year. Under the multinational nuclear deal negotiated with Iran in 2015 by former U.S. President Barack Obama's administration, Iran agreed to convert the reactor to run on low-enriched uranium fuel, which would produce only traces of plutonium. Two years ago, the United States announced it was withdrawing from the deal. Last week, U.S. Secretary of State Mike Pompeo cited "Iran's continued nuclear escalation" as the basis for ending the waivers, which also covered reducing proliferation risk at the Tehran Research Reactor. Iranian officials say they will continue to reconfigure Arak on their own.

ITER builders take key step

FUSION ENERGY | Construction crews last week lifted into place the first major piece of the $25 billion ITER project, which will be the world's largest fusion reactor and may finally demonstrate that melding together hydrogen nuclei is a viable energy source. In last week's step, about 200 workers carefully lifted the 1250-ton, 30-meter-wide steel base of the reactor's containment chamber into the construction pit near Cadarache in France. The COVID-19 pandemic has slowed progress by delaying delivery of some components; if key ones are delayed past December 2021, it may become difficult to complete construction by 2025 as planned, says ITER Director-General Bernard Bigot.



Cranes lift the steel base of ITER's containment chamber, called a cryostat, into place.

PHOTO: ITER ORGANIZATION

U.S. limits Chinese visas

POLICY | Chinese graduate students with ties to institutions in China that do military research can no longer receive a visa to study in the United States under an executive order issued last week by President Donald Trump. Those already in the United States with such ties could have their immigration status reviewed. The new policy is part of an effort by the Trump administration and Republicans in Congress to block what they see as China's widespread theft of U.S.-funded academic research. Higher education officials worry such steps—including another pending bill targeting Chinese financial support of U.S.-based researchers—will limit the pool of foreign talent.

Published by AAAS

Men, like this one being carried into a medical center for COVID-19 patients in Moscow, tend to get sicker from the coronavirus than women.

IN DEPTH

COVID-19

Sex hormones signal why virus hits men harder

Emerging role of androgens suggests potential therapies to suppress infection

By **Meredith Wadman**

In January, one of the first publications on those sickened by the novel coronavirus in Wuhan, China, reported that three out of every four hospitalized patients were male. Data from around the world have since confirmed that men face a greater risk of severe illness and death from COVID-19 than women, and that children are largely spared. Now, scientists investigating how the virus does its deadly work have zeroed in on a possible reason: Androgens—male hormones such as testosterone—appear to boost the virus' ability to get inside cells.

A constellation of emerging data supports this idea, including COVID-19 outcomes in men with prostate cancer and lab studies of how androgens regulate key genes. And preliminary observations from Spain suggest that a disproportionate number of men with male pattern baldness—which is linked to a powerful androgen—end up in hospitals with COVID-19. Researchers are rushing to test already approved drugs that block androgens' effects, deploying them early in infection in hopes of slowing the virus and buying time for the immune system to beat it back.

"Everybody is chasing a link between androgens … and the outcome of COVID-19," says Howard Soule, executive vice president at the Prostate Cancer Foundation, who on 13 May ran a Zoom call presenting the newest research that drew 600 scientists and physicians. A second call scheduled for 3 June will discuss incipient clinical trials.

Epidemiological data from around the world have confirmed the early reports of male vulnerability. In Lombardy in Italy, for example, men comprised 82% of 1591 patients admitted to intensive care units (ICUs) from 20 February to 18 March, according to a *JAMA* paper. And male mortality exceeded that of women in every adult age group in another *JAMA* study of 5700 New York City patients hospitalized with COVID-19.

Now, researchers are on the trail of a mechanism for this male bias—an effort led by prostate cancer researchers, who have a deep acquaintance with androgens. Christina Jamieson of the University of California (UC), San Diego, who has developed organoids to study prostate cancer, recalls that she was in a Zoom meeting honing ideas on how to link her research to COVID-19 when her sister, also a UC San Diego scientist, sent her a one-word text. It read: "TMPRSS2."

It was 16 April, and within minutes Jamieson had found the publication that prompted the text: a *Cell* paper by Markus Hoffmann of the Leibniz Institute for Primate Research and colleagues. The paper sent a lightning bolt through the prostate research community, because it showed that infection with SARS-CoV-2, the virus that causes COVID-19, relies in part on TMPRSS2, a membrane-bound enzyme. The enzyme cleaves the "spike" protein on the coronavirus' surface, allowing the virus to fuse with the host cell's membrane and get inside the cell.

Jamieson and other prostate cancer researchers were familiar with the enzyme, because in about half of all prostate cancers, a TMPRSS2 mutation revs up an oncogene that kicks cell growth into overdrive. In the prostate, TMPRSS2 is produced when male hormones bind to the androgen receptor. "Doing research, it's like you're trying to throw an anchor into the vast ocean of

A protective treatment?

In one Italian study, men with prostate cancer who received drugs that suppress androgens were much less likely to be infected with SARS-CoV-2.

	MEN ON ANDROGEN-DEPRIVATION THERAPY (ADT)	MEN NOT ON ADT
Total men with prostate cancer	5273	37,161
Number infected with SARS-CoV-2	4	114
Estimated cases per 10,000	8	31

CREDITS: (PHOTO) VYACHESLAV PROKOFYEV/TASS VIA GETTY IMAGES; (DATA) MONTOPOLI *ET AL.*, *ANNALS OF ONCOLOGY*, HTTPS://DOI.ORG/10.1016/J.ANNONC.2020.04.479 (2020)

Published by AAAS

possibilities," Jamieson says. The discovery that TMPRSS2 helps the virus enter cells "felt like the anchor hit ground."

Researchers haven't established that androgens control TMPRSS2 in the lung—ground zero for SARS-CoV-2 infection—as they do in the prostate; studies in lung tissue and cells from mice and humans come to conflicting conclusions. But after the *Cell* paper was published, Andrea Alimonti, head of molecular oncology at Università della Svizzera italiana, strengthened the androgen link by looking at data on more than 42,000 men with prostate cancer in Veneto in Italy. He and colleagues found that patients on androgen-deprivation therapy (ADT)—drugs that slash levels of testosterone—were only one-quarter as likely to contract COVID-19 as men with prostate cancer not on ADT, they reported in the *Annals of Oncology* (see table, p. 1038). Men on ADT were also less likely to be hospitalized and to die, although numbers were small.

Another retrospective study, still unpublished, controlled for age and other medical conditions and got similar results: Of 58 patients with prostate cancer who contracted the coronavirus, the 22 taking ADT were significantly less likely to be hospitalized and to need supplemental oxygen, says William Oh, a prostate cancer physician-scientist at the Icahn School of Medicine at Mount Sinai. "Our conclusion supports the hypotheses that androgen signaling might increase the risk of severe outcomes from COVID-19 and that androgen deprivation may limit those severe outcomes," Oh says.

Science's COVID-19 coverage is supported by the Pulitzer Center.

Two small studies have reported that men with male pattern baldness are overrepresented among hospitalized COVID-19 patients. This type of baldness is associated with high levels of dihydrotestosterone (DHT), a key metabolite of testosterone, in the scalp. An April study of 41 Spanish men hospitalized for COVID-19 found that 71% had male pattern baldness; the background rate in white men is estimated at 31% to 53%. A second study published last month found that 79% of 122 men in three Madrid hospitals with COVID-19 had male pattern baldness.

More circumstantial evidence comes from stem cell biologist Faranak Fattahi, of UC San Francisco. Her team found a strong link between a measure of active androgens in the blood and the severity of COVID-19 disease in data from several hundred male patients in the UK Biobank; they did not find this effect in women.

Such evidence is already inspiring possible therapies. Matthew Rettig, an oncologist who directs prostate cancer research at UC Los Angeles, is leading a double-blind, randomized, placebo-controlled trial of the androgen-suppressing drug degarelix in 200 veterans hospitalized with COVID-19 in Los Angeles, Seattle, and New York City. Patients in the active arm will receive a single injection that virtually zeroes out testosterone levels within 3 days. That reduces expression of the *TMPRSS2* gene, at least in the prostate, to almost nil. Side effects include hot flashes and breast growth and "are equivalent to surgical castration," Rettig says.

But whereas in prostate cancer, the injections are given month after month, "This study only involves a one-time dosage. It's temporary," Rettig says. He hopes to learn in 4 to 5 months whether the treatment helps keep patients off ventilators and reduces mortality.

Several other antiandrogen trials are in the offing in the United States and Europe. Prostate cancer researcher Catherine Marshall of Johns Hopkins University is preparing a trial of bicalutamide, an older, inexpensive androgen receptor blocker, in 20 patients hospitalized within 3 days after they tested positive for COVID-19. Her group will compare outcomes with patients who don't receive the drug. "We think that if this works it's going to work by decreasing the viral load in patients," Marshall says. "That's why we are doing it earlier in people's course of disease."

Women are being included in the trial, she adds, because they have androgens, although at lower levels than men, and because estrogens have been shown to help heal acute lung injury. Bicalutamide raises estrogen levels as well as suppressing androgen activity. Marshall says of the emerging wave of trials: "All these trial ideas have been team science at its best and probably at its fastest."

Adding to the promise of antiandrogens is lab-based evidence from Fattahi's study. Her team screened Food and Drug Administration (FDA)-approved drugs in heart cells in the lab to see which ones reduced levels of the essential SARS-CoV-2 receptor, angiotensin-converting enzyme 2 (ACE2). Key hits included finasteride and dutasteride, drugs that block the conversion of testosterone to DHT, according to a 15 May preprint. Finasteride is FDA-approved to treat male pattern baldness and dutasteride for prostate enlargement. Dutasteride also reduced ACE2 levels in healthy human lung alveolar cells.

Although researchers pursuing the androgen link caution that their hypothesis remains just that until it is borne out in lab and clinical studies, they're optimistic. "When all evidence points to the same thing it's very satisfying," Fattahi says. ■

COVID-19

Blood vessel injury may spur disease's fatal second phase

Damage to vessel lining may drive mysterious clotting disorders, inflammation

By **Catherine Matacic**

Frank Ruschitzka told his pathologist to be ready before the first COVID-19 patient died. In early March, Ruschitzka, who leads the cardiology department at University Hospital Zürich, noticed that patients with the disease had strange symptoms for what was then thought to be chiefly a respiratory infection. Many patients had acute kidney failure, organ damage, and mysterious blood clots. Several weeks later, the first body was autopsied: Tiny clots and dead cells littered the capillaries of the lungs, and inflammation had distended blood vessels supplying every organ in the body.

The pathologist had never seen anything like it. But the results showed Ruschitzka why his patients were suffering so much: The virus had targeted their blood vessels.

Since the Zürich team's findings were published in mid-April, dozens of studies have revealed similar patterns of vascular damage in people who died of COVID-19. For example, a 21 May paper in *The New England Journal of Medicine* showed that the lungs of COVID-19 victims had nine times as many clots as those who died of the H1N1 flu. Other studies have noted inflammatory symptoms in children (*Science*, 29 May, p. 923) and strokes in otherwise healthy young adults. Now, researchers have woven these findings into a new hypothesis explaining why some patients slip into a fatal "second phase" of COVID-19, 1 week or so after hospitalization.

The key is direct and indirect damage to the endothelial cells that line the blood vessels, particularly in the lungs, explains Peter Carmeliet, a vascular biologist at the Belgian research institute VIB and co-author of a 21 May paper in *Nature Reviews Immunology*. By attacking those cells, COVID-19 infection causes vessels to leak and blood to clot. Those changes in turn spark inflam-

Published by AAAS

mation throughout the body and fuel the acute respiratory distress syndrome (ARDS) responsible for most patient deaths.

"It's a vicious cycle," says Nilam Mangalmurti, a pulmonary intensivist at the Hospital of the University of Pennsylvania, who was not involved in the new research.

This mechanism could explain why the disease pummels some patients who have obesity, diabetes, and cardiovascular conditions: The cells lining their blood vessels are already compromised. If so, drugs used to treat these conditions might help prevent other COVID-19 patients from sliding into serious disease. "[A vaccine] would be terrific," says Richard Becker, a cardiologist at the University of Cincinnati College of Medicine who outlined a similar cardiovascular cascade in a 15 May review in the *Journal of Thrombosis and Thrombolytis*. But until a safe, effective vaccine is available, he says, such therapeutics might be "a good start."

In healthy individuals, endothelial cells help regulate blood pressure, prevent inflammation, and inhibit clotting, in part through the continual production of nitric oxide (NO); they also serve as gatekeepers for molecules passing in and out of the bloodstream. When injured, they send out a complex array of signals to immune cells and clotting factors, which rush to repair the site. And they warn their fellow endothelial cells to be on alert for invaders.

Based on autopsy reports like those from the Zürich hospital, the epidemiology of the disease, and how the new coronavirus behaves in cells in the lab, Carmeliet and colleagues believe the virus can send that system spinning out of control.

When SARS-CoV-2 enters the lungs, it invades cells in the air sacs that transfer oxygen to the blood. Surrounding those sacs are capillaries lined like bricks with endothelial cells. The virus directly invades some of those cells; others become "activated," likely in response to signals from the invading virus and other damaged cells. Some infected cells likely commit suicide. "It's not a quiet death where the cell just dies," Mangalmurti says. "All the contents leak out."

Carmeliet and colleagues suggest damage and other changes in the activated cells trigger vascular leakage, flooding the air sacs with fluid, a hallmark of ARDS. White blood cells swarm to the lungs and NO production likely plummets. Together with the activated endothelial cells, the immune cells release a host of signaling molecules, including interleukins, which raise local blood pressure and weaken cell junctions. Damage to the endothelial cells also exposes the membrane underneath them.

That exposed membrane in turn triggers uncontrolled clotting. The endothelial and immune cells add fuel to the fire, recruiting additional clotting factors and platelets, which help form clots. Those clots degrade into the key biomarker D-dimer, creating the sky-high levels that alert clinicians to patients in trouble (see graphic, below). Eventually, such clotting spreads throughout the body and blocks the blood supply within vital organs.

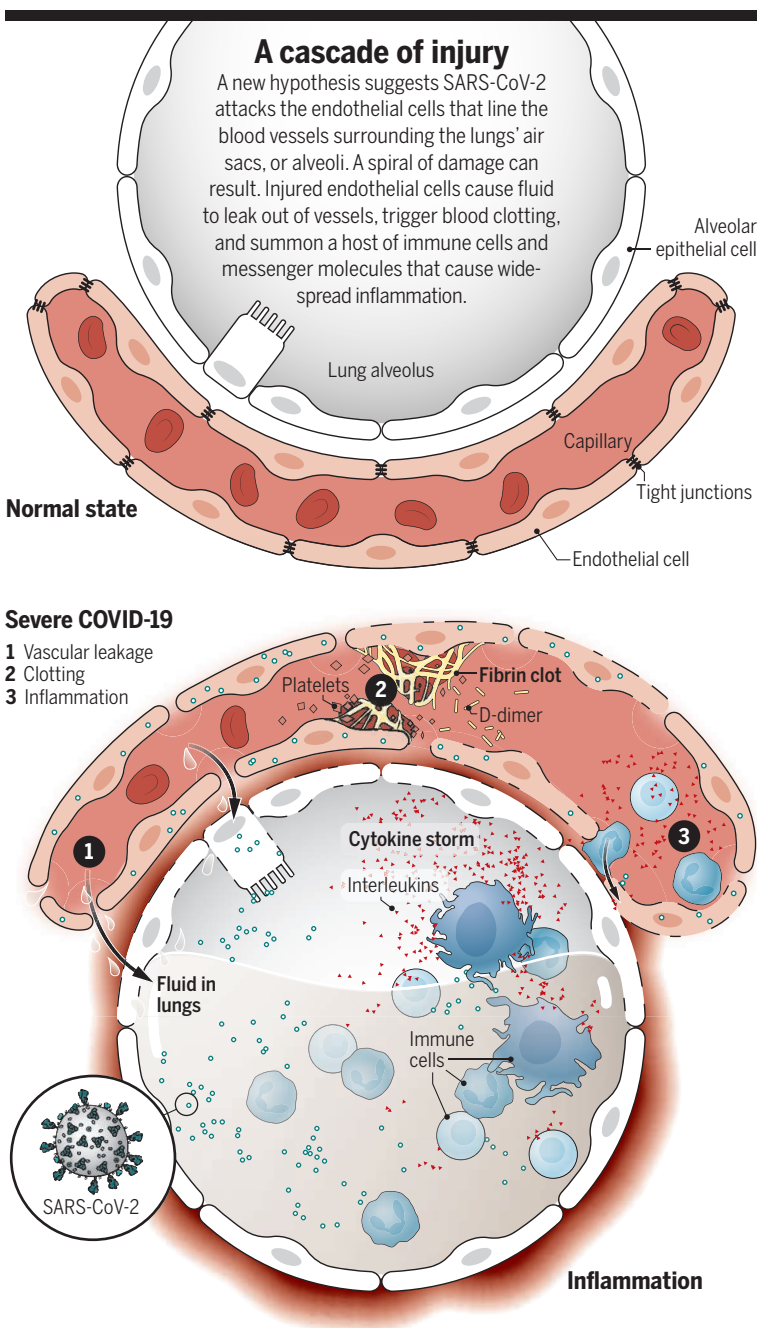
These chain reactions culminate in a final, destructive phase of inflammation. Like clotting, inflammation is an essential defense, sending a diverse army of cells and messenger molecules called cytokines to fight invaders and mop up the debris of battle. But in COVID-19, this reaction spirals out of control in a deadly cytokine storm and plunges patients' bodies into shock.

Ruschitzka says the three-step hypothesis "makes perfect sense" of what he saw in his patients; he's already sending the Carmeliet paper to colleagues. He says the array of pathways may also explain why some young people without known risk factors for COVID-19 become seriously ill: They might have undiagnosed clotting or autoimmune disorders, such as rheumatoid arthritis, that amplify the effects of SARS-CoV-2 infection.

This emerging view of the key role of endothelial cells suggests that a number of existing drugs might dampen or even arrest the fatal second phase of the disease, Becker says. Already, evidence that inflammation and clotting play a role in COVID-19 has inspired dozens of trials in the United States and Europe of anticlotting, anti-inflammatory, and antiplatelet drugs.

Ruschitzka thinks another commonly prescribed drug might help: statins. Typically taken to lower cholesterol, they also reduce inflammation and improve endothelial cell function.

Mangalmurti welcomes such trials, but cautions that patients may respond differently depending on how healthy their endothelial cells are to start. "One size does not fit all." ■



GRAPHIC: TEUWEN *ET AL., NAT. REV. IMMUNOL.* (2020); ADAPTED BY V. ALTOUNIAN/*SCIENCE*

Published by AAAS

Corrected 4 June 2020. See full text.

A hydroxychloroquine study is being audited.

PHOTO: AP PHOTO/JOHN LOCHER

COVID-19

The pandemic's first major research scandal erupts

Critics question patient data used to challenge malaria drugs

By **Kelly Servick** *and* **Martin Enserink**

On its face, it was a major finding: Antimalarial drugs touted by the White House as possible COVID-19 treatments looked to be not just ineffective, but downright deadly. A study published on 22 May in *The Lancet* used hospital records procured by a little-known data analytics company called Surgisphere to conclude that COVID-19 patients taking chloroquine or hydroxychloroquine were more likely to show an irregular heart rhythm—a known side effect thought to be rare—and more likely to die. Within days, large randomized trials of the drugs screeched to a halt. Solidarity, the World Health Organization's (WHO's) megatrial of potential COVID-19 treatments, paused recruitment into its hydroxychloroquine arm.

But just as quickly, the results have begun to unravel—and Surgisphere, which provided patient data for two other high-profile COVID-19 papers, has come under withering online scrutiny from researchers and amateur sleuths. They have pointed out many red flags in the *Lancet* paper, including the astonishing number of patients and details about patient demographics and dosing that seemed implausible. "It began to stretch and stretch and stretch credulity," says Nicholas White, a malaria researcher at Mahidol University in Bangkok.

As *Science* went to press, *The Lancet* issued an Expression of Concern, noting "serious scientific questions" about its paper. Hours earlier, *The New England Journal of Medicine* (*NEJM*) issued an Expression of Concern about a second study using Surgisphere data, published on 1 May. The paper reported that taking certain blood pressure drugs including angiotensin-converting enzyme (ACE) inhibitors didn't increase the risk of death among COVID-19 patients, as some researchers had suggested. The journal asked the authors "to provide evidence that the data are reliable."

A third study using Surgisphere data is also under fire. In an April preprint, Surgisphere founder and CEO Sapan Desai and coauthors concluded that ivermectin, an antiparasitic drug, dramatically reduced mortality in COVID-19 patients. In Latin America, where ivermectin is widely available, that study led some officials to authorize use of the drug, creating a surge in demand.

Chicago-based Surgisphere has not publicly released data underlying the studies. On 2 June, Desai told *Science* through a spokesperson that he was "arranging a nondisclosure agreement that will provide the authors of the *NEJM* paper with the data access requested by *NEJM*." And in a 29 May statement, Surgisphere defended the integrity of its research and said it was pursuing "an independent academic audit" of its results in *The Lancet*. The journal and non-Surgisphere authors also said data reviews were underway.

The episode has left leaders of halted hydroxycholoroquine trials weighing whether to restart. "The problem is, we are left with all the damage that has been done," says White, a co-investigator on a halted trial for COVID-19 prevention. It will now be hard to recruit people to key studies, he says. "The whole world thinks now that these drugs are poisonous."

Desai co-authored the *Lancet* paper with cardiologist Mandeep Mehra of Harvard University's Brigham and Women's Hospital (BWH), cardiologist Frank Ruschitzka of University Hospital Zürich, and cardiac surgeon Amit Patel, who listed affiliations with the University of Utah and HCA Research Institute in Nashville, Tennessee. (Mehra and Patel referred inquiries to BWH. Ruschitzka did not respond to requests for comments.) The authors describe an analysis of electronic health record data from patients already treated for COVID-19 at 671 hospitals on six continents—nearly 15,000 people prescribed chloroquine or hydroxychloroquine, alone or in combination with an antibiotic, and a control group of 81,000 other patients. After adjusting for potentially confounding factors, the researchers found the risk of dying was 9.3% for the control group versus 23.8% for those getting hydroxychloroquine alongside an antibiotic.

In a 25 May media briefing, WHO Director-General Tedros Adhanom Ghebreyesus cited the results in announcing a "temporary pause" in Solidarity's hydroxychloroquine arm. Regulators in France and the United Kingdom also instructed investigators, including White's team, to halt enrollment in trials. And Sanofi said it would temporarily stop recruiting patients to two trials of its hydroxychloroquine formulation.

Other researchers immediately took issue with the analysis. The study does not properly control for the likelihood that patients getting the experimental drugs were sicker than the controls, says Matthew Semler, a critical care physician at Vanderbilt University. And White notes anomalies in the data. Although 66% of the patients were reportedly treated in North America, the reported doses tended to be higher than the guidelines set by the U.S. Food and Drug Administration. And the authors claim to have included 4402 patients in Africa, but it seems unlikely that African hospitals would have detailed electronic health records for so many patients, White says. The

Published by AAAS

Corrected 4 June 2020. See full text.

study also reported more deaths in Australian hospitals than the country's official COVID-19 death statistics, *The Guardian* reported. On 29 May, *The Lancet* issued a correction saying a hospital assigned to the study's "Australasia" group should have been assigned to Asia and updating a supplemental table. "There have been no changes to the findings of the paper," it says.

The brief response left some researchers frustrated. "This was very, very annoying," says James Watson, a statistician at Mahidol who on 28 May published an open letter—now signed by more than 140 researchers—that calls for the release of Surgisphere's hospital-level data, an independent validation of the results, and publication of the peer-review comments that led to the *Lancet* publication. "*The Lancet* encourages scientific debate and will publish responses to the study, along with a response from the authors," a journal spokesperson said in a response.

On 2 June, many of the same researchers and others published an open letter to *NEJM* and the authors of the ACE inhibitor study, citing similar problems in that paper. It notes inconsistencies including a discrepancy between the small number of hospitals in each country that are said to have shared patient data with Surgisphere and the high proportion of those countries' confirmed COVID-19 cases included in the study.

Oddities also appear in the ivermectin study, says Carlos Chaccour of the Barcelona Institute for Global Health. There's evidence that ivermectin, the key weapon in the global campaign against river blindness, also has antiviral properties. The 6 April preprint, co-authored by Patel, Desai, and Mehra, along with David Grainger of the University of Utah, used Surgisphere data reportedly collected at 169 hospitals around the world between 1 January and 1 March. It included three patients in Africa who received ivermectin—even though only two COVID-19 cases had been reported in all of Africa by 1 March, Chaccour and two colleagues note in a recent blog post.

Chaccour says after he inquired about the discrepancy, the authors posted a second, longer version of the manuscript on 19 April, containing data collected between 1 January and 31 March. The new manuscript reported that ivermectin reduced the need for mechanical ventilation by 65% and slashed the death rate by 83%. But the revision had other problems, Chaccour and his colleagues wrote in their blog post. For example, the data shown in a figure were wildly different from those reported in the text. (Grainger also did not reply to a request for a comment.)

In response to the ivermectin study the Peruvian Ministry of Health modified its COVID-19 treatment protocol to include ivermectin (as well as hydroxychloroquine) for mild and severe cases of COVID-19; demand for the drug in Peru has surged. In Trinidad, Bolivia, the city government aimed to hand out more than 350,000 free doses of ivermectin after the country's Ministry of Health authorized its use against COVID-19.

Surgisphere's sparse online presence—the website doesn't list partner hospitals by name or identify its scientific advisory board, for example—has prompted intense skepticism. Physician and entrepreneur James Todaro of the investment fund Blocktown Capital wondered in a blog post why Surgisphere's enormous database doesn't appear to have been used in peer-reviewed research studies until May. Chaccour asks how such a tiny company—LinkedIn lists only a handful of employees—was able to reach data-sharing agreements with hundreds of hospitals around the world.

Desai's spokesperson says the company has 11 employees and has been developing its database since 2008.

The potential of hydroxychloroquine for treating COVID-19 has become a political flashpoint. French microbiologist Didier Raoult, whose own widely criticized studies suggested a benefit from the drug, derided the *Lancet* study in a video posted on 2 June, calling the authors "incompetent."

For scientists running randomized trials of hydroxychloroquine, an urgent question has been how to respond to the paper and the ensuing flap. A trial funded by the U.S. National Heart, Lung, and Blood Institute opted to keep running after its data and safety monitoring board (DSMB) reviewed safety data from already enrolled participants, says Semler, a co-investigator on the study. WHO's paused Solidarity trial is awaiting similar review from its DSMB, says Soumya Swaminathan, the organization's chief scientist.

The controversy is an unfortunate distraction, says Miguel Hernán, a Harvard epidemiologist and co-investigator on an ongoing trial of hydroxychloroquine in Spain and Latin America. "If you do something as inflammatory as this without a solid foundation, you are going to make a lot of people waste time trying to understand what is going on." Chaccour says both *NEJM* and *The Lancet* should have scrutinized the provenance of Surgisphere's data more closely before publishing the studies. "Here we are in the middle of a pandemic with hundreds of thousands of deaths, and the two most prestigious medical journals have failed us," he says. ■

With reporting by Rodrigo Pérez Ortega, Charles Piller, and John Travis.

COVID-19

Shuttered natural history museums fight for survival

But creative scientists vow recovery and move research and public programs online

By **Elizabeth Pennisi**

A few months ago, retirement was the furthest thing from David Thomas's mind. "Then the world went upside down," recalls the archaeologist from the American Museum of Natural History in New York City. In March, the coronavirus pandemic forced the museum to close its doors. No more school groups thronging the interactive exhibits, no more corporate dinners or lines of international tourists waiting to pay $23 a head to marvel at fossils. The museum's income plummeted 60%.

Leaders first asked for early retirements. By early May, they had sliced the staff of 1100 by 20% and furloughed an additional 250 staff members. Many full-time employees now work 3 days a week, mostly from home. Thomas opted to retire early, along with four of the other 38 curators. "It was the right thing to do," he says.

Around the world, natural history museums are shuttered and reeling. Last week, the California Academy of Sciences announced it was furloughing or laying off 40% of its staff. "We will recover, but there is no doubt that we will be in some ways a different institution," says Peter Roopnarine, a paleontologist there.

Museums' reliance on revenue from ticket sales and events makes them among the first scientific institutions to feel the economic impact of the COVID-19 pandemic. "I worry about the long-term health of all natural history museums and the collections that are in our sacred trust," says Shannon Hackett, an ornithologist at the Field Museum of Natural History in Chicago. "It will be very challenging for some museums to reopen at all," adds Scott Cooper, who runs Drexel University's Academy of Natural Sciences in Philadelphia.

But the crisis is also spurring museums

Published by AAAS

The Field Museum of Natural History hosted a socially distanced blood drive in its empty halls.

PHOTO: SCOTT OLSON/GETTY IMAGES

to adopt or expand practices that, though they may not restore lost revenue, are keeping the public engaged and research ticking along: an online biodiversity contest, public discussions on Zoom, a webcam streaming captive corals. Curators are also expanding and refining digital collections that are accessible to both the public and homebound researchers.

"We are seeing more changes in the museum industry at this moment than we could push people to make previously," says Julie Stein, director of the Burke Museum of Natural History and Culture at the University of Washington, Seattle, whose own institution has been devastated. The Burke Museum had opened a new building in October 2019 and was "headed for record-breaking revenue," Stein says—until the entire campus shut down on 6 March.

Some museums, including the Smithsonian Institution's National Museum of Natural History in Washington, D.C., have dodged financial cliffs thanks to government support. The Natural History Museum, London, stayed afloat with emergency support from the U.K. government, but furloughed half its staff until the end of June. Similarly, the Field Museum has thus far avoided layoffs thanks to a cash reserve and the federal paycheck protection program, says President and CEO Richard Lariviere. But 30% of the museum's income comes from tickets and related activities, and $17 million is already lost. Given that cases of COVID-19 have yet to peak in Illinois, Lariviere doubts the museum will open this summer, and worries he will be forced to make layoffs.

Some university museums managed to avoid layoffs now, but may pay a price later if university budgets shrink. Harvard University's Peabody Museum of Archaeology and Ethnology will likely not reopen as quickly as stand-alone museums, says Director Jane Pickering.

As they worry about the future, researchers are also distraught because they can't pursue their current research. Travel restrictions have brought fieldwork to a screeching halt—and with it, the addition of more specimens to collections. The American Museum of Natural History alone has canceled 100 expeditions. And researchers can't get into buildings to analyze existing collections. "We have been cut off from our collections, facilities, and colleagues," says Anjali Goswami, a paleobiologist at the Natural History Museum, London.

One trend accelerated by the crisis could help: efforts to digitize natural history collections. At Harvard's Museum of Comparative Zoology, staff working from home have been busy enhancing the millions of records in the museumwide database, for example adding latitude and longitude coordinates to specimens thus far identified only by location names. "There's a tremendous amount of data locked into collections," says Kirk Johnson, director of the National Museum of Natural History, where detailed digital images of pressed plants in the herbarium allow researchers to scrutinize them from afar. "We are now shining the light on the dark data of museums."

Rebecca Albright, a coral biologist at the California Academy of Sciences, is studying the mysteries of coral spawning; only one other research team has been able to get the corals to reproduce in a lab setting. Recently, Albright identified just the right conditions, including water temperature and lighting that re-create changing day length and the cycling of the Moon, to prompt spawning. When she learned that she couldn't be in the museum at the key time, she and her colleagues set up an infrared webcam. "We never set up a camera before because we didn't need to," she says.

The livestreaming camera allowed them to catch spawning in the act on 22 April, Earth Day—and made a big splash on the web. "If we had missed this, we would have had to wait a whole year," Albright says. The corals now have 1.6 million followers.

Other scientists have refocused their research on the pandemic itself. Roopnarine previously studied how nature recovered from mass extinctions. Now, he is repurposing his computer models of ecosystem recovery to evaluate how various employment schemes may get economies back on track as lockdowns ease. "Our work has never been more relevant than it is now," he says.

Many see the pandemic as an opportunity for change. "I'm doing more public programs than ever—in virtual formats," says Sabrina Sholts, a biological anthropologist at the National Museum of Natural History.

To engage the public in research, the Natural History Museum of Los Angeles County and the California Academy of Sciences in April enlisted thousands of citizen scientists in a global biodiversity effort called the City Nature Challenge. Participants gathered thousands of images of birds, insects, and other wildlife in more than 250 cities to help researchers study urban ecosystems.

The pandemic itself is inspiring new directions. Leonard Krishtalka, director of the University of Kansas Biodiversity Institute and Natural History Museum, wants museums to expand their focus to include microbes and viruses.

At the National Museum of Natural History, Sholts was already thinking along those lines when she pulled together a temporary exhibit inspired by the 2014 Ebola epidemic. "Outbreak," which explores how pathogens spread between animals and people, opened 2 years ago. After COVID-19 erupted, the exhibit was replicated online, with a digital version in half a dozen languages; a DIY kit has already been adapted in 41 countries and 30 U.S. states and territories.

"We are in the process of reinventing what natural history museums are for," Johnson says, speaking by phone to a reporter as he walked past the darkened halls of the Outbreak exhibit. "Museums can play a much more impactful role than they have in the past 50 years." ■

Published by AAAS

European Commission President Ursula von der Leyen has proposed a €1.85 trillion budget that bets on science.

FUNDING

Europe bets R&D spending will bring jobs to battered economy

Horizon Europe gets €13.5 billion to spend fast, spur growth

By **Nicholas Wallace**

The European Union wants a massive dose of research spending to lift it out of what could be the worst recession in its history. Last week, as part of a €1.85 trillion budget and pandemic recovery proposal, the European Commission, the EU executive arm, unveiled plans to pump €94.4 billion into research over 7 years, nearly €11 billion more than originally planned for the program, called Horizon Europe. But not everyone thinks the money is the best medicine.

The Commission says R&D spending will drive productivity, employment, and competitiveness. Based on its own models, it has estimated that Horizon 2020, the €80 billion predecessor to Horizon Europe, will ultimately create up to 179,000 jobs and add €400 billion to €600 billion to gross domestic product (GDP) by 2030. Likewise, the new infusion "will ensure that we come out stronger from this crisis," said EU research commissioner Mariya Gabriel in a statement.

But Julia Lane, an economist at New York University who studies the effects of R&D spending, says Europe is not tracking the data needed to justify its predictions and, in any case, its numbers aren't credible. If the return on R&D investment for GDP is that good, "why don't they just shut the entire economy down and move it into that?" she asks. "Just on the face of it, it's a ludicrous number."

The Horizon Europe budget still has to be negotiated with national governments. But the Commission wants €13.5 billion of the research money to be spent fast, within 4 years, on projects in four broad categories: health; climate, energy, and mobility; digital, industry, and space; and the new European Innovation Council (EIC). The Commission will define specific topics and call for grant proposals from researchers; it will also consider making equity investments in startups through EIC (*Science*, 10 April, p. 120).

In the short term, research to curtail the pandemic will do the most to help the economy, says Thomas Estermann, head of funding policy at the European University Association, a lobby group. "Each week of confinement and lockdown has a tremendous economic impact," he says. "The solution to the crisis will clearly come from research." The money won't be available until January 2021, however, and only a tiny portion of the €13.5 billion is likely to fund COVID-19 treatments or vaccines.

In the long term, the value of R&D spending is plain to everyone, says Robert-Jan Smits, president of the Eindhoven University of Technology and former head of the Commission's research policy department. "If I ask a person here in the streets, 'What are the sources that bring us prosperity and wealth?' he or she will say, 'education and technology,'" he says. "It's not even rocket science."

While at the Commission, Smits brought in economists to develop one of the models that predicts economic outcomes based on historical data. The model, called NEMESIS, "can calculate for each investment, each euro that you invest in science and innovation at EU level, what will be the impact—even on jobs, even on growth," he says.

But without detailed accounting from research centers, it is impossible to make reasonable estimates of job creation, says Lane, who in a former role at the U.S. National Science Foundation helped develop metrics to track job creation from the Obama administration's 2009 stimulus package, which put billions into research. She says setting up the infrastructure to do that in the United States took years.

According to Lane, the U.S. data show some impact: For example, researchers are more likely to found successful startups, which contribute disproportionately to job creation and productivity. But more than 10 years later it still isn't clear what kinds of science funding work best, she says. Europe, meanwhile, has "wasted the last 10 years" by failing to set up similar procedures, she says.

Smits agrees that tracking the immediate impact of research investment is "a bit complex." But he argues that identifying longer term impacts in models is "doable," because NEMESIS, unlike conventional models, factors in R&D spending. "I find it just completely bizarre that people still work in the finance ministries with models which are not giving any recognition to key technologies, which are changing our lives on a daily basis," Smits says.

Hans-Olaf Henkel, a former member of the European Parliament and former president of the Leibniz Association, an association of German research institutes, objects to the Commission's top-down approach to stimulating the economy. "Any kind of centralized Brussels program, in my view, is doomed to fail," he says. He argues politicians don't know what kind of science is most likely to succeed, and should therefore concentrate on bottom-up funding. "They should let the scientists do their creative work and not come up with political goals." The €13 billion European Research Council, one of the few parts of Horizon 2020 that focuses on bottom-up funding and is meant to be immune from political interference, was originally slated to get a €3.6 billion rise in Horizon Europe, but the reallocation of money under the new budget puts that figure in doubt.

Regardless of how the new budget shakes out, Andrés Rodríguez-Pose, an economic geographer at the London School of Economics, points out that EU grants are "a drop in the sea." Most R&D funding in Europe comes from the private sector, and most of the rest comes from national governments, not the European Union. "The capacity of the EU institutions to shape the way research is done in Europe is limited," he says. ■

Nicholas Wallace is a journalist in Brussels.

PHOTO: REUTERS/JOHANNA GERON

Published by AAAS

U.S. SCIENCE POLICY

Bill would supersize NSF's budget—and role

Legislation calls for $100 billion increase, new technology directorate, and new name

By **Jeffrey Mervis**

The National Science Foundation (NSF) would get a huge infusion of cash, as well as a new name and new responsibilities for keeping the United States on top in technological innovation, under bipartisan bills introduced late last month in both chambers of Congress. But some science policy veterans are questioning whether a basic research agency should also be expected to spearhead the development of new technologies.

The Endless Frontier Act (S. 3832 and H.R. 6978) would create a technology directorate at NSF with a budget that could grow to $35 billion by 2024—more than four times the agency's existing $8 billion budget. That would bring NSF to rough parity with the National Institutes of Health, whose $41 billion budget makes it by far the government's biggest funder of basic research.

The bill would authorize spending $110 billion over 5 years, with $100 billion for NSF (see chart, right) and $10 billion for the Department of Commerce to set up a dozen or so regional technology hubs to promote innovation in areas not currently tech hot spots. NSF would also get a second deputy director to oversee all technology activities and a new name: the National Science and Technology Foundation.

The funding levels are aspirational; Congress would still have to appropriate the money even if the bill were adopted. But academic leaders view the legislation as a huge vote of confidence in NSF, which turned 70 this year. "These funds—which complement, not supplant, existing resources, an important condition—build on the NSF's strengths and would fill gaps in our research enterprise," says Rafael Reif, president of the Massachusetts Institute of Technology.

The legislation addresses the chronic underfunding of NSF, says Neal Lane, who spent 5 years as NSF director before becoming science adviser to former President Bill Clinton. "NSF gets enough good ideas to justify" a much larger budget, he says. "This bill makes clear that it's time for such bold action."

But Arden Bement, who led NSF under former President George W. Bush, says other federal agencies already have the mission of supporting applied technology and development. And no federal agency, he adds, will ever have enough resources to substitute for what industry spends on commercializing innovation.

The bill's name invokes the title of the seminal 1945 report by presidential science adviser Vannevar Bush that made the case for federal support of academic research and led to NSF's creation in 1950. But the Stay Ahead of China Act might be a more accurate moniker, based on what lawmakers say they hope it will accomplish.

"China and others are stealing American intellectual property and aggressively investing in research and commercialization to dominate the known technology fields of the future," the four co-sponsors—Senators Charles Schumer (D–NY) and Todd Young (R–IN) and Representatives Ro Khanna (D–CA) and Mike Gallagher (R–WI)—say in the preface to their bills. "The country that wins the race in key technologies—such as artificial intelligence, quantum computing, advanced communications, and advanced manufacturing—will be the superpower of the future," they add.

The legislation calls on NSF to fund an unspecified number of university-based technology centers in those and other fields. The centers would be an order of magnitude bigger than NSF's existing engineering and science centers, which have annual budgets of up to $5 million. Partnering with industry scientists, the centers would both carry out fundamental research and develop prototypes of high-tech products and processes. NSF would also receive billions to expand its six science directorates, boost investment in education and training, and set up facilities to test new technologies.

The legislation could significantly alter how NSF operates, with the Defense Advanced Research Projects Agency (DARPA) as a model. It allows NSF to adopt DARPA practices, including fixed-term appointments of experts from the private sector and a focus on tangible, deadline-driven results. "The new [technology] directorate can run like DARPA if NSF wants it to," says one university lobbyist familiar with Schumer's thinking.

Because NSF currently gives grantees freedom to pursue curiosity-driven research, adopting a DARPA-like model "would be a huge cultural shift," says Joel Parriott, director of public policy for the American Astronomical Society and a former White House budget official whose portfolio included NSF. "It's also not clear how the technology directorate could operate so differently from the rest of the agency."

The bill's prospects this year are iffy in a Congress consumed by the coronavirus pandemic and with scant time to legislate before the fall elections. One option is attaching it to a reauthorization of defense programs that is seen as must-pass legislation. But in the meantime, David Hart, a science policy expert at George Mason University, says its "symbolic value" shouldn't be ignored.

"It's a bipartisan statement that the country is underinvesting in key technologies," Hart says. "I'm not sold on doing all of this at NSF. But we, as a nation, have to come up with new ways to fund technology. This is certainly a breathtaking proposal."

Bement disagrees. "Action on this bill should be tabled for another day," he says. Instead, he suggests Congress first "determine whether the system is broken and, if so, in what ways." ■

Dividing up the tech dollars

Lawmakers have proposed spending $110 billion over 5 years to boost U.S. innovation. The Commerce Department would receive $10 billion to establish up to 15 regional technology hubs with industry partners, whereas the National Science Foundation (NSF) would get $100 billion, distributed across these six categories:

University-based technology centers	35%	For basic research, prototyping, and support of regional hubs
NSF priorities	20%	Allocated as needed within NSF and to other U.S. agencies
Education and training	15%	For scholarships, fellowships, and traineeships
Research, including social and ethical concerns	15%	Money channeled through various NSF programs
Testbeds and fabrication	10%	Support for facilities and processes to shorten time to market
Fostering entrepreneurship	5%	Strengthening the innovation enterprise in academia

DATA: U.S. CONGRESS

Published by AAAS

FEATURES

DOUBLE TROUBLE

The brightest supernovae, used as cosmic yardsticks, have an unexpected trigger: pairs of dead stars

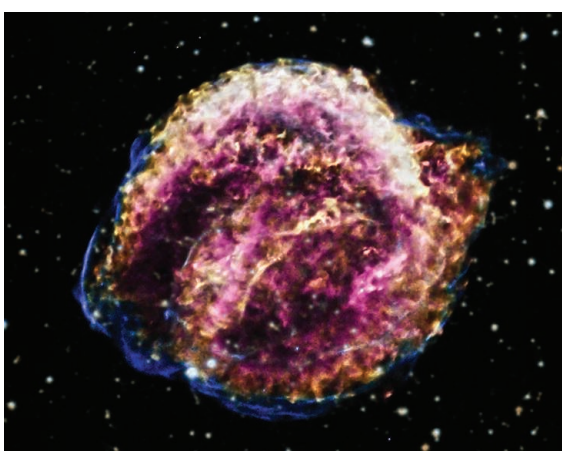
By **Daniel Clery**

The white dwarf stars were zipping across the Milky Way at more than 1000 kilometers per second—thousands of times faster than a speeding bullet, so fast that they would eventually escape the gravitational clutches of the Galaxy. "They were not like anything we had seen before," says Boris Gaensicke, an astronomer at the University of Warwick.

Gaensicke and his colleagues suspected these burnt-out embers were fleeing scenes of violence: supernova explosions in which another white dwarf had detonated like an Earth-size hydrogen bomb. In the standard picture of these explosions, known as type Ia supernovae, a nearby giant star lights the fuse. But the extreme speed of the white dwarfs suggested a different scenario, in which the fleeing dwarfs had delivered the sparks, from close orbits around the doomed stars. When they blew up, these partners were flung away like shots from a biblical sling.

The speeds of the three white dwarfs, discovered in a 2018 data set from Europe's Gaia satellite, were just one clue to this picture. Subsequent ground-based observations found traces of iron and other metals in the stars' light—elements that might have been implanted by a supernova blast. From the color and brightness of the light, the astronomers could deduce that the stars were hotter and larger than typical white dwarfs, as if they had been puffed up by a sudden energy boost. Most telling, the researchers rewound the stars' trajectories and found that one hailed from a known site: the remnant of a 90,000-year-old supernova. They are "the very best evidence" for the twin white dwarf scenario, says team leader Ken Shen of the University of California (UC), Berkeley.

The evidence that twin white dwarfs drive most, if not all, type Ia supernovae, which account for about 20% of the supernova blasts in the Milky Way, "is more and more overwhelming," says Dan Maoz, director of Tel Aviv University's Wise Observatory, which tracks fast-changing phenomena such as supernovae. He says the classic scenario of a white dwarf paired with a large star such as a red giant "doesn't happen in nature, or quite rarely."



Kepler's supernova, a type Ia explosion from 1604, lingers today with the x-ray glow of hot, leftover debris.

Which picture prevails has impacts across astronomy: Type Ia supernovae play a vital role in cosmic chemical manufacturing, forging in their fireballs most of the iron and other metals that pervade the universe. The explosions also serve as "standard candles," assumed to shine with a predictable brightness. Their brightness as seen from Earth provides a cosmic yardstick, used among other things to discover "dark energy," the unknown force that is accelerating the expansion of the universe. If type Ia supernovae originate as paired white dwarfs, their brightness might not be as consistent as was thought—and they might be less reliable as standard candles.

Robert Kirshner, a longtime supernova watcher at the Gordon and Betty Moore Foundation, isn't ready to give up on the classic scenario, but he acknowledges the misgivings about it. "It's plausible that there is more than one path to a type Ia supernova, but the uniformity of the light output is a tiny bit of a paradox," he says. Nevertheless, "There's a nagging doubt: Do we fully understand the nature of these explosions?"

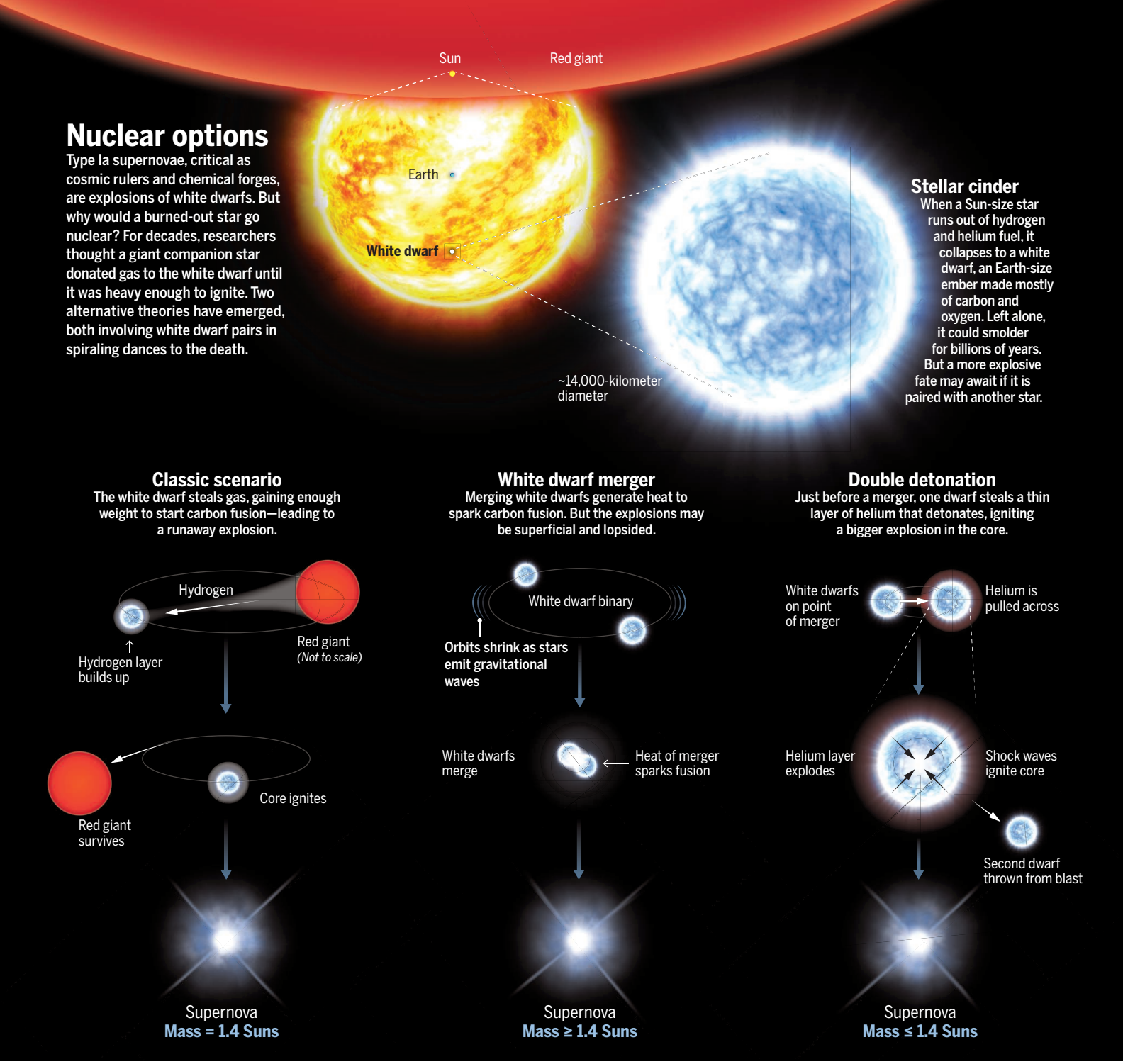
SUPERNOVAE COME in two basic flavors. The most common are called core-collapse. They occur when a massive star, far larger than the Sun, runs out of fuel. The pressure of the core's heat can no longer counteract gravity; the outer parts of the star collapse into the core with such force that a rebounding shock wave, or sometimes a renewed fusion burn, blows out the star's outer layers, leaving a neutron star or black hole behind.

The rest are all type Ia—white dwarf stars somehow reignited into a runaway fusion reaction. The burst of new energy happens too fast for the star to absorb, blowing the entire thing to smithereens in a blast that is brighter and longer than a core-collapse supernova.

White dwarfs might seem unlikely candidates for fireworks. They are the cinders of Sun-like stars that have burned up their hydrogen and helium fuels, leaving—in most cases—carbon and oxygen, heavier elements that can't fuse in such a low-mass star. White dwarfs shrink to the size of Earth, and only glow from leftover heat. Theoretically, they

CREDITS: (X-RAY) NASA/CXC/SAO/D. PATNAUDE; (OPTICAL) DSS

Published by AAAS



GRAPHIC: C. BICKEL/SCIENCE

should cool to black over billions of years.

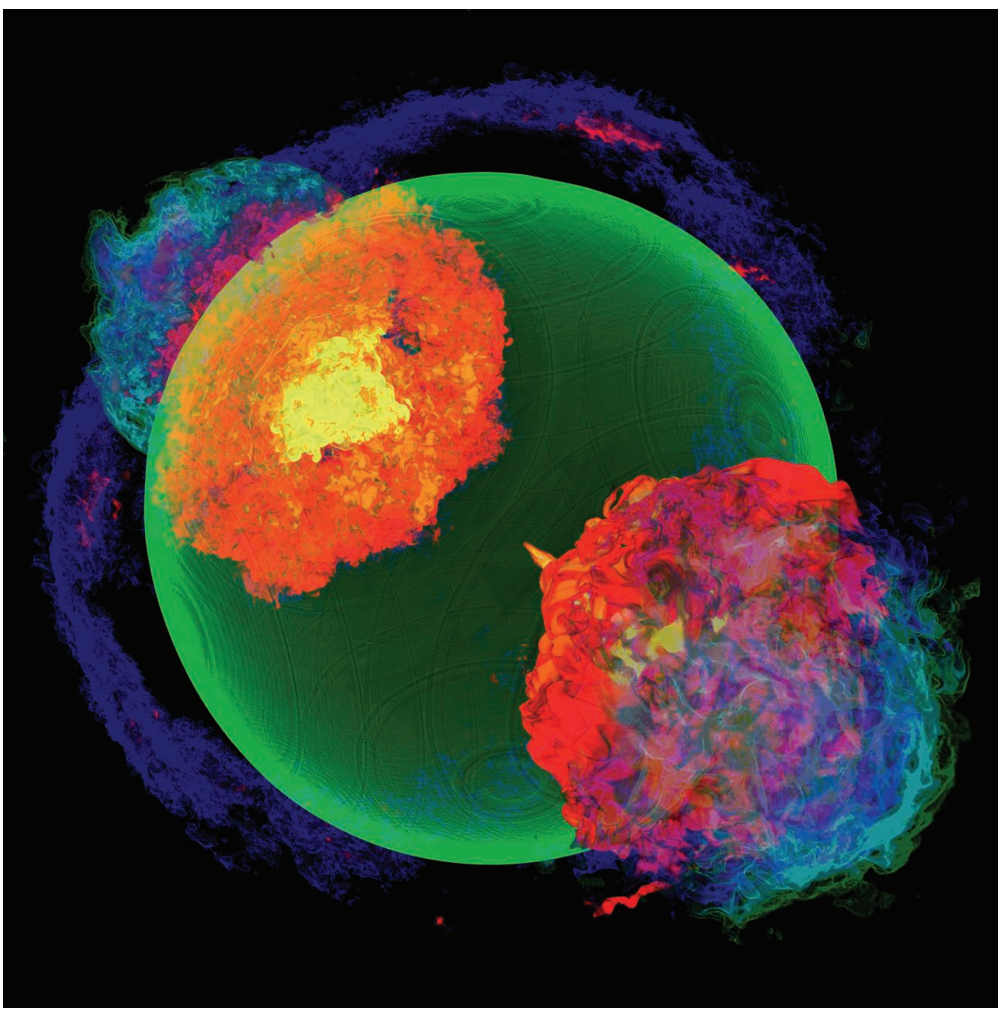
But if the white dwarf orbits in a binary pair with another star, a more spectacular fate may await it. The classic type Ia scenario was proposed in 1973 by John Whelan and Icko Iben. They mapped out the fading light of supernova SN 1972e for one full year, and realized its brightness could be explained by the decay of about one solar mass worth of radioactive metals. These, they proposed, were forged in a white dwarf that had grown to a size at which the pressure and temperature in its core would be high enough to fuse carbon, causing a thermonuclear blast. Whelan and Iben suggested the white dwarf might grow to that mass threshold if its gravity stole hydrogen gas from a companion star such as a puffed-up red giant, which doesn't clutch its outer layers too tightly.

Later modeling showed achieving this growth was a tricky balancing act. If a white dwarf gobbles up hydrogen too fast, the hydrogen layer that forms on its surface can get hot enough to blow up prematurely, in a more modest thermonuclear explosion called a classical nova or, if it happens repeatedly, a recurrent nova. If it accumulates too slowly, the white dwarf can tiptoe up to a mass 1.44 times that of the Sun. At that threshold, known as the Chandrasekhar limit, theorists predict the pressure inside will cause electrons and protons to fuse into neutrons, and the white dwarf will quietly collapse into a neutron star.

But if the hydrogen is added at just the right rate, a white dwarf, especially one rich in carbon and oxygen, can respond

Published by AAAS

more dramatically. Just short of the Chandrasekhar limit, at about 1.4 solar masses, the density and temperature of the core shoot up. Flares of carbon fusion break out. After smoldering for 100 or more years, a runaway reaction detonates the star and blows it apart in a matter of seconds. The resulting fireball, 5 billion times as bright as the Sun, forges a suite of metals from chromium to nickel in the periodic table. The radioactive decay of nickel to cobalt and then iron powers a brilliant afterglow that peaks in a couple of weeks and fades over years. And because, in this picture, every type Ia explodes with the same mass, they should all have the same peak brightness.



Theorists have struggled to simulate classic type Ia explosions. This is one of their successful models.

THAT SCENARIO satisfied astronomers for decades. But they have yet to find definitive evidence for it. After a supernova disperses enough to be transparent, astronomers routinely search for a surviving red giant companion but have never found one.

Models also suggest a flash of blue light should appear, hours after the supernova begins, as the expanding fireball slams into the tenuous hydrogen atmosphere of the red giant and heats it enough to glow in the ultraviolet. But when astronomers spotted a supernova in the nearby Pinwheel galaxy within hours of its ignition in 2011, they saw no blue flash. "SN 2011fe was a paradigm changer," says Stan Woosley, a supernova expert at UC Santa Cruz.

A second predicted signal—weaker but more persistent—has also been elusive. Astronomers would expect some hydrogen from the red giant to be swept along with the other explosion debris, creating an absorption line in the spectrum of the supernova's light as the remnant cools. But a 2019 study of 227 type Ia supernova remnants found no hint of this hydrogen.

The numbers are against the classic scenario, Maoz says. In a galaxy like the Milky Way, a type Ia supernova occurs once every few hundred years. If they all originate from a white dwarf and a red giant, the galaxy would have to host some 10,000 of these pairs—and there's little evidence of them. Because it's hard to see binary pairs as separate stars at such great distances, astronomers have looked for indirect evidence, such as the recurrent novae triggered when hydrogen from a red giant spills quickly onto a white dwarf companion, or the soft x-ray glow that can result from a steady stream of hydrogen. If the Milky Way harbored 10,000 white dwarf-red giant pairs, we should see many novae and soft x-ray sources, but only a handful have been found, Maoz says. "We're missing 99.9% of the systems."

The theoretical underpinnings of the classic scenario have problems, too. A key assumption is that the explosion is so violent that the entire star is blown apart. But theorists have had a hard time modeling how the smoldering fusion flares would burgeon into an all-consuming detonation. Modeling that process requires simulating nuclear reactions with centimeter resolution across an object the size of Earth for 100 years or more. Until recently, computers couldn't handle that challenge—and now that they can, Shen says, only some models predict the transition. "The state of the field is not conclusive yet."

ENTER ANOTHER WAY to blow up a white dwarf. Astronomers know that two white dwarfs orbiting each other—a so-called double degenerate—can gradually spiral inward as their rapid orbits throw out gravitational waves, which carry away energy. When they merge, there would be easily enough heat and pressure to kick-start carbon fusion in the combined stars.

Models suggest a problem with this mechanism: The burning would not start in the core. It might kick off in the hot zone where the two are actively colliding, leading to a lopsided and incomplete explosion that would leave much of the stars' mass unburned. Or it could ignite close to the surface, only blowing off the outer layers of the merged star. "Getting a detonation in a double degenerate is by no means trivial," Woosley says.

Yet Maoz says the odds are irresistible: White dwarf binaries are thought to

IMAGE: FLASH CENTER FOR COMPUTATIONAL SCIENCE/UNIVERSITY OF CHICAGO

Published by AAAS

be fairly common. In a 2018 study, Maoz and his colleagues looked for the wobble of white dwarfs being tugged by a partner, combining survey results from the Sloan Digital Sky Survey and Europe's Very Large Telescope. They concluded that about half a billion white dwarf binaries have merged in the Milky Way since its formation. If just one-sixth of those mergers led to a type Ia supernova, that would be one supernova every 200 years—roughly what is observed in the Milky Way. Another team, using Gaia data, came to a similar conclusion.

Maoz is undeterred by the problem of getting a complete, symmetrical blast. "Just because we don't know how it happens doesn't mean nature hasn't found a way." In fact, many believe nature has found a way to blow up one member of a white dwarf pair—but without a merger. White dwarfs can have some leftover helium in their atmospheres after the core stops burning. When an orbiting pair is on the cusp of merging, the larger of the two stars can rapidly steal helium from the smaller one to form a dense helium layer on its surface. The helium layer can act as a kind of blasting cap, exploding in a small thermonuclear blast and sending a shock wave into the star that can ignite the core.

This scenario is called D6, for dynamically driven double-degenerate double detonation. The idea was first developed in 2010 by James Guillochon, a researcher at the Harvard-Smithsonian Center for Astrophysics, and his colleagues. It leaves the smaller white dwarf battered but thrown free, like those Shen's group found in the Gaia data. D6 was originally thought to require a hefty amount of helium, making it a rare event, but more recent modeling suggests just a few percent of a solar mass could be enough, Woosley says.

ONE KEY FEATURE of the D6 scenario is that the exploding white dwarf can be well below the critical mass, because the spark comes from a shock wave and not from gravitational pressure. A less massive exploding star will produce less nickel and be less bright.

Recent studies of the metallic elements supernovae forge suggest the low-mass type Ia may be the norm. According to models, the production of manganese in type Ia supernovae is particularly sensitive to density in the white dwarf's core: If the star is close to the 1.4–solar mass threshold, its high-density core produces lots of manganese; if the star is lighter—as is likely in the D6 mechanism—it produces one-tenth as much.

As a result, manganese abundances derived from the light of stars today can hint at the masses of the ancient supernovae that seeded them with heavier elements. "Manganese provides an indirect way to probe previous generations of type Ia supernovae that went off in that galaxy," says Ashley Ruiter of the University of New South Wales, Canberra.

In a pioneering study from 2013, researchers led by Ivo Seitenzahl, then at the Julius Maximilian University of Würzburg, compared manganese abundance in the Sun with models of how much manganese would be produced by supernovae of different masses. They found that only half of the supernovae that exploded in the solar neighborhood in the past needed to be high mass to explain the Sun's manganese content. "This was the first of a new wave of results," says Maria Bergemann of the Max Planck Institute for Astronomy. This year, she and her colleagues reported looking at manganese in 42 stars across the Milky Way and concluded that the abundances suggest 75% of the galaxy's type Ia supernovae were low mass.

> **"It's a real golden age for supernovae because we're finding so many."**
>
> **Andy Howell**,
> Las Cumbres Observatory

The implications of undersize type Ia supernovae extend far beyond the elements in the present-day universe. They also raise questions about the explosions' longstanding role as "standard candles" for probing cosmic history.

In 1998, researchers compared a few dozen distant type Ia supernovae with closer ones and found that they were dimmer than they should have been. They concluded that the universe's expansion is accelerating, driven by some unknown dark energy—a discovery for which they were awarded the 2011 Nobel Prize in Physics. Supernova distances are also at the heart of a dispute over the value of the expansion rate itself, known as the Hubble constant (*Science*, 10 March 2017, p. 1010). In the nearby universe the expansion rate is measured using standard candles such as type Ia supernovae; in the distant, early universe, it is derived from clues such as the cosmic microwave background, the echo of the big bang. When the effect of dark energy is taken out, the two values should agree—but they don't.

Could a not-so-standard candle jeopardize those discoveries? "It means something, but not that dark energy goes away," Woosley says. Dark energy has been confirmed using other methods, so he's not worried about that. But he thinks cosmologists will run into trouble as they put their theories to more rigorous tests that require more precise standard candles. "Supernovae could be less useful for precision cosmology," he says.

Astronomers already knew the peak brightness of type Ia supernovae isn't perfectly consistent. To cope, they have worked out an empirical formula, known as the Phillips relation, that links peak brightness to the rate at which the light fades: Flashes that decay slowly are overall brighter than those that fade quickly. But more than 30% of type Ia supernovae stray far from the Phillips relation. Perhaps low-mass D6 explosions can explain these oddballs, Shen says. For now, those who wield the cosmic yardstick will need to "throw away anything that looks weird," Gaensicke says, and hope for the best.

Andy Howell, a supernova watcher at Las Cumbres Observatory, thinks type Ia supernovae could still be reliable tools for cosmology if astronomers could separate the different varieties of type Ia that are now lumped together. "If we knew there were two populations, we could make the measurements even better," he says.

So far, astronomers can't say how many of their favorite cosmic explosions are sparked by white dwarf pairs rather than a giant and a dwarf. "It's too early to say with certainty what that fraction is," Ruiter says. But the coming years could bring more clarity.

Survey telescopes that scan the skies nightly or even hourly are catching more and more supernovae. The current frontrunner, the Zwicky Transient Facility in California, spots about 30 supernovae per night. Its output will be dwarfed in 2022 with the opening of the Vera C. Rubin Observatory, an 8.4-meter survey telescope in Chile that is expected to find thousands of supernovae nightly. Other telescopes able to obtain spectra from thousands of objects simultaneously will enable astronomers to study the explosions for the features—the blue flash, the hydrogen absorption lines—that could betray the involvement of a giant star.

Shen and Gaensicke hope the next data release from Gaia will contain more high-speed white dwarfs fleeing from D6 explosions. And the Laser Interferometer Space Antenna, an orbiting gravitational wave detector due for launch in 2034, will be able to sense white dwarf pairs as they spiral in toward merger, giving astronomers a better idea of how common they really are. "It's a real golden age for supernovae because we're finding so many," Howell says. "We've now finally got the tools to see them in new ways." ■

Published by AAAS

ECOLOGY

Blue carbon from the past forecasts the future

Coastal mangrove ecosystems must rise above sea level to survive

By **Catherine E. Lovelock**

When knowledge of a system is imperfect, thresholds are extremely useful tools for decision-makers. Sea level rise (SLR) threatens intertidal coastal wetlands because these aquatic plants will drown as water depths exceed their physiological tolerance limits. For mangroves—trees and shrubs that grow in tidal waters of the tropics or subtropics—the threshold for SLR has been elusive. On page 1118 of this issue, Saintilan *et al.* (*1*) deduce that the threshold SLR for mangrove ecosystems is 6 to 7 mm/year. This discovery can help inform decisions on how to sustain mangroves, which provide 200 million coastal people with essential ecosystem services. These include protection from intense storms and waves, reduction in the impact of coastal flooding, sequestering of carbon, improvement in water quality, and preservation of biodiversity and fisheries (*2*).

PHOTO: JURGEN FREUND/MINDEN PICTURES

Published by AAAS

Sea level rise threatens coastal mangrove forests. Disruption of these ecosystems endangers coastal communities and decreases carbon sequestration.

Saintilan *et al*. studied ancient mangrove sediments to infer a threshold of relative SLR for mangrove tolerance. The term "relative" is used because the rate of SLR is determined not only by the increase in water volume of the oceans but also by subsidence or uplift of coastal land. Subsidence and uplift result, in part, from glacial isostatic adjustment, which describes a complex set of geophysical processes that vary geographically. The authors used sediment cores to collate the dates when mangrove vegetation first appeared in a variety of locations over the past 10,000 years. Mangroves can be detected from their distinctive, highly organic sediments, which reflect mangrove ecosystems' great productivity and conditions that favor preservation of vast quantities of organic carbon. The data obtained from sediment cores, sampled by a range of researchers from 78 sites on five continents, indicate that mangrove ecosystems did not develop unless relative SLR was less than 6 to 7 mm/year.

The global mean rate of SLR is now 3.4 mm/year and is projected to exceed the threshold and reach ~10 mm/year by 2100 under "business-as-usual" scenarios (greenhouse gas concentration trajectory Representative Concentration Pathway 8.5) (*3*). However, SLR could stabilize at ~5 mm/year by the year 2100 under moderate-emissions scenarios in which countries reduce carbon dioxide (CO_2) emissions (*3*). Clearly, reducing CO_2 in the atmosphere is a first line of defense against passing the relative SLR threshold for mangrove persistence. However, in many mangrove locations, rates of relative SLR are already higher than 6 to 7 mm/year. Now, in the Anthropocene, subsidence of coastal land is also caused by extraction of oil, gas, and water (*4*). For example, the Mekong Delta of Vietnam is subsiding at a rate of 6 to 20 mm/year (*5*) and the Ganges-Brahmaputra Delta by 1 to 7 mm/year (*6*), which is accompanied by land erosion and saltwater intrusion. At the same time, sediment supply to the coast, which is vital to counteracting the effects of relative SLR as it contributes to raising the seafloor, has declined; this is because rivers are dammed and, in some cases, sediment has been mined and exported, all of which further increase the vulnerability of mangroves to SLR (*7*).

Saintilan *et al*. also found variation around the mean threshold value of 6 to 7 mm/year, indicating that local factors, such as subsidence and sediment supply, drive location-specific relative SLR thresholds for mangroves. Reduction in extraction of water, oil, and gas from floodplains and enhancing sediment supply might avoid crossing the SLR threshold for mangrove persistence. Dam removal and the delivery of nourishing sediments and water to the coast are practiced in Europe and North America but have little uptake in the tropics (*8*). Improvement of water quality and reduction of overexploitation will help maintain mangrove health by facilitating strong root growth, which also contributes to vertical accretion (*9*). Balancing the freshwater and sediment needs of mangroves that provide natural ecosystem services against water storage, irrigated agriculture, and hydropower is an emerging series of trade-offs that policymakers must face (*10*).

One key consideration in these trade-offs is that under moderate SLR, mangroves mitigate CO_2 in the atmosphere by storing it in the plants and sediment of the ecosystem (so-called blue carbon) (*11*). Saintilan *et al*. used their sediment core data to estimate the amount of CO_2 sequestered over ~1400 years of mangrove development and reached a conservative estimate of 85 Pg of stored carbon, which is larger than the estimated carbon stored in boreal peats as they developed through the same Holocene period. These historical estimates verify that sequestration

School of Biological Sciences, The University of Queensland, St Lucia, QLD 4072, Australia. Email: c.lovelock@uq.edu.au

Published by AAAS

of blue carbon in coastal ecosystems makes substantial contributions to the mitigation of global climate change (*12*).

Without sufficient sediment and root growth to reduce the impacts of SLR, mangrove cover might be sustained through landward migration of mangroves onto floodplains. This would increase carbon sequestration, but landward retreat is complex and can be costly for governments and landholders (*13*). The threshold provided by Saintilan *et al.*, combined with knowledge of local factors that influence relative SLR, provides opportunities to develop time lines for decisions and actions that will maintain mangroves in the land-seascape.

There are caveats to consider in the new study. Over the late Holocene, rates of SLR were mostly declining, whereas Earth currently faces accelerating SLR, which might alter the threshold. There could be lags in responses of mangrove cover to relative SLR that exceed the 6- to 7-mm/year threshold, and lags can vary with species traits and disturbance regimens, such as frequency and intensity of storms that exacerbate the impacts of SLR and which are predicted to intensify with climate change in many regions.

Nevertheless, providing an evidence-based relative SLR threshold for mangrove survival can help stimulate solutions for coastal management. If nations and communities wish to harness the potential of blue carbon to mitigate climate change and to protect millions of people who depend on mangroves for shelter, flood protection, food, and fiber, then solutions for staying below the 6- to 7-mm/year relative SLR threshold should be the goal of civil society and governments. With reduction in CO_2 emissions, equitable management of river water and sediment flows, limitations in water and oil extraction on floodplains, and conservation and planning for landward migration, mangroves can be maintained into the future. ■

REFERENCES AND NOTES

1. N. Saintilan *et al.*, *Science* **368**, 1118 (2020).
2. E. B. Barbier *et al.*, *Ecol. Monogr.* **81**, 169 (2011).
3. Intergovernmental Panel on Climate Change (IPCC), "Climate Change 2013: The Physical Science Basis. Working Group 1 Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change," T. F. Stocker *et al.*, Eds. (IPCC, 2013).
4. J. P. Syvitski *et al.*, *Nat. Geosci.* **2**, 681 (2009).
5. P. S. J. Minderhoud *et al.*, *Sci. Total Environ.* **634**, 715 (2018).
6. M. Becker *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 1867 (2020).
7. C. E. Lovelock *et al.*, *Nature* **526**, 559 (2015).
8. L. Ding *et al.*, *Chin. Geogr. Sci.* **29**, 1 (2019).
9. K. W. Krauss *et al.*, *New Phytol.* **202**, 19 (2014).
10. G. Grill *et al.*, *Nature* **569**, 215 (2019).
11. E. Mcleod *et al.*, *Front. Ecol. Environ.* **9**, 552 (2011).
12. K. Rogers *et al.*, *Nature* **567**, 91 (2019).
13. C. R. Field *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 9134 (2017).

10.1126/science.abc3735

IMMUNOLOGY

The specifics of innate immune memory

Innate immune cells develop specific memory involved in transplant rejection in mice

By **Jorge Domínguez-Andrés**[1] and **Mihai G. Netea**[1,2]

One of the most important traits of immune host defense against pathogens is memory, which improves survival if the same pathogen is reencountered. However, immune memory can also be deleterious, driving autoimmune diseases and the rejection of transplanted organs. Memory characteristics have been considered a fundamental property of adaptive immune cells such as T and B lymphocytes (*1*). However, innate immune cells such as myeloid cells and natural killer (NK) cells can also adapt to previous encounters with pathogens through epigenetic, transcriptional, and functional reprogramming, called trained immunity (*2*). The discovery of this innate immune memory emerged from studies with live vaccines and was described as being largely nonspecific (*3*). On page 1122 of this issue, Dai *et al.* (*4*) reveal that monocytes and macrophages acquire specific memory and induce organ rejection in mice, which could be prevented to improve transplantation outcomes.

Dai *et al.* use a series of elegant transplant models in different mouse strains, showing that the rejection of grafts is maintained in individuals without a functional adaptive immune response. Graft rejection was mediated by monocytes and macrophages, which retained their memory functions several weeks after the encounter with antigens from the donor. This specific memory response was dependent on paired immunoglobulin-like receptor-A (PIR-A), an activation receptor expressed by B cells and myeloid lineage cells, which can recognize major histocompatibility complex class I (MHC-I) molecules (see the figure). MHC-I expressed on the surface of most nucleated cells presents peptide fragments of endogenous proteins, which can be subsequently recognized as antigens by the immune cells of the recipient of a transplant and cause its rejection (*5*). Blocking or deleting PIR-A receptors impaired the recognition of MHC-I–peptide complexes expressed in the donor cells, which decreased the rejection of transplanted hearts and kidneys and improved outcomes.

This specificity of innate immune memory in mouse myeloid cells described by Dai *et al.* is reminiscent of memory characteristics of innate immune cells of several invertebrates, which lack an adaptive immune system. These include diversification of genes encoding fibrinogen-related proteins (FREPs) in mollusks (*6*) and scavenger receptor cysteine-rich proteins in echinoderms (*7*) or alternative splicing of the immunoglobulin (Ig) domain–encoding gene Down syndrome cell adhesion molecule (*Dscam*) in insects (*5*). It is important to note that in both vertebrates and invertebrates, the Ig superfamily of molecules is often the pillar mediating specificity of the innate immune memory responses. Indeed, PIR-A, B cell receptors, immunoglobulins, and T cell receptors share a similar type of structure, which argues for an evolutionary continuum of memory specificity in innate and adaptive immune responses.

Ig-based specific immune responses complement the more primitive nonspecific trained immunity-mediated memory (*8*). The increased responsiveness provided by trained immunity after certain infections or vaccinations can induce protection against both specific and heterologous infections (*9*, *10*). Yet, in conditions characterized by excessive immune responses, such as inflammatory and autoimmune diseases and organ transplant rejection, enhanced innate immune responses can aggravate the pathological consequences of inflammation. It is crucial to know the roles played by innate immune cells in transplant rejection, in order to target it and improve survival.

The findings of Dai *et al.* have several implications that go beyond transplantation. If monocytes and macrophages can develop specific memory to MHC-I–presented antigens, this reveals possible therapeutic approaches in organ transplantation, but also autoimmune and inflammatory diseases. In addition, it is intriguing to hypothesize whether specific innate immune memory

[1]Department of Internal Medicine and Radboud Center for Infectious Diseases, Radboud, Netherlands. [2]Immunology and Metabolism, Life and Medical Sciences Institute, University of Bonn, Bonn, Germany. Email: mihai.netea@radboudumc.nl

Published by AAAS

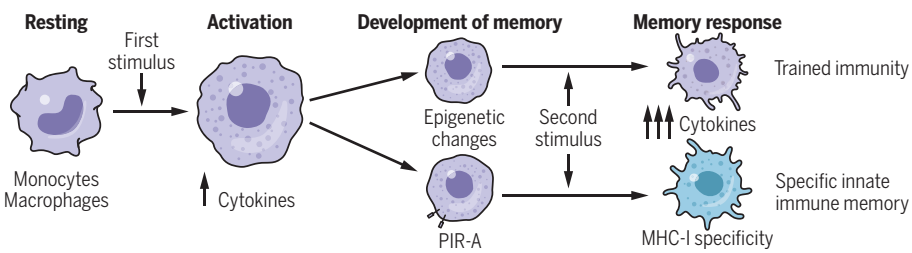
responses could also form against pathogens, as proposed for NK cells in response to viral infections (*2*).

Macrophages are heterogeneous tissue-resident cells: Yolk sac– and fetal liver monocyte–derived tissue-resident macrophages colonize different tissues during embryonic development (*11*). These are long-lived and able to self-renew, so the specific characteristics of these macrophages rely on their niche (*11*). For example, lung macrophages are exposed to airway antigens, whereas Kupffer cells in the liver are exposed to gut-derived molecules (*12*). The phenotype of a microglial cell in the brain greatly differs from that of a peritoneal macrophage (*11*). The potential ability of these and other myeloid cell subsets to develop different types of immunological memory to antigens could offer additional evidence for their different functions across tissues and help to explain the development of different types of memory responses to the same antigens in different locations. In this context, it will be crucial to address if the antigen-specific responses by myeloid cells can also be inhibitory and may be involved in the development of immunological tolerance leading to the lack of responsiveness against harmless molecules such as antigens expressed by commensal bacteria, or self-antigens. Immunological tolerance prevents harmful immune responses in the host, whereas failure of these mechanisms results in tissue damage and autoimmune responses (*13*). Indeed, polymorphisms in genes of the human leukocyte antigen (HLA) system, the human version of MHC, are associated with celiac disease, inflammatory bowel disease, rheumatoid arthritis, psoriasis, type 1 diabetes, and multiple sclerosis (*14*, *15*). Additional studies have found genetic associations between MHC-I and infectious diseases such as HIV, human hepatitis B virus, and tuberculosis (*14*).

If monocytes and macrophages develop antigen-specific memory to antigens presented by donor-derived MHC-I molecules, it is likely that these and other myeloid cells can develop immunological memory to self- and nonself antigens of different sources. Given the variety of receptors expressed by the different types of myeloid cells, the implications of these mechanisms could encompass many processes beyond organ transplantation, including the response to pathogens, vaccines, inflammatory and autoimmune diseases, or the development of allergies. Dai *et al.* identified several families of polymorphic Ig superfamily receptors that could bind to MHC-I molecules. The extension of this search to other receptors that can bind antigens of a different nature could lead to the identification of other structures with the potential to trigger or block antigen-specific memory in myeloid cells, potentially offering a new set of targets for immunotherapy.

Different forms of innate immunological memory

Some myeloid cells can develop memory. After a first stimulus, retained epigenetic changes can facilitate the production of cytokines after a nonspecific stimulus (trained immunity). Macrophages can also develop specific memory through paired immunoglobulin-like receptor-A (PIR-A) specificity to antigens presented by major histocompatibility complex class I (MHC-I).



GRAPHIC: JOSHUA BIRD/*SCIENCE*

REFERENCES AND NOTES

1. F. A. Bonilla, H. C. Oettgen, *J. Allergy Clin. Immunol.* **125** (suppl. 2), S33 (2010).
2. M. G. Netea *et al.*, *Nat. Rev. Immunol.* 10.1038/s41577-020-0285-6 (2020).
3. L. C. J. de Bree *et al.*, *Semin. Immunol.* **39**, 35 (2018).
4. H. Dai *et al.*, *Science* **368**, 1122 (2020).
5. F. L. Watson *et al.*, *Science* **309**, 1874 (2005).
6. S. M. Zhang, C. M. Adema, T. B. Kepler, E. S. Loker, *Science* **305**, 251 (2004).
7. L. Du Pasquier, *Science* **309**, 1826 (2005).
8. M. G. Netea, A. Schlitzer, K. Placek, L. A. B. Joosten, J. L. Schultze, *Cell Host Microbe* **25**, 13 (2019).
9. S.-C. Cheng *et al.*, *Science* **345**, 1250684 (2014).
10. E. Kaufmann *et al.*, *Cell* **172**, 176 (2018).
11. M. Guilliams, G. R. Thierry, J. Bonnardel, M. Bajenoff, *Immunity* **52**, 434 (2020).
12. A. J. Macpherson, M. Heikenwalder, S. C. Ganal-Vonarburg, *Cell Host Microbe* **20**, 561 (2016).
13. J. A. Bluestone, *Immunol. Rev.* **241**, 5 (2011).
14. V. Matzaraki, V. Kumar, C. Wijmenga, A. Zhernakova, *Genome Biol.* **18**, 76 (2017).
15. J. Domínguez-Andrés, M. G. Netea, *Trends Immunol.* **40**, 1105 (2019).

ACKNOWLEDGMENTS

M.G.N. is supported by a European Research Council Advanced Grant (no. 833247) and a Spinoza Grant.

10.1126/science.abc2660

DEVELOPMENT

Testicular-borne factors affect sperm fertility

Immature testicular germ cells secrete factors that influence sperm maturation in mice

By **Tessa Lord**[1] *and* **Jon M. Oatley**[2]

Male fertility relies on the genesis of sperm in seminiferous tubules of the testis and their maturation during transit through the epididymis. Mouse models with impaired development of the most proximal region of the epididymis, the initial segment (IS), possess sperm that are morphologically normal but incapable of fertilizing an egg (*1*). It has been postulated that factors synthesized in the testis are released into the lumen of tubules (lumicrine factors) and influence development and function of the IS. However, the identity of such factors has remained elusive. On page 1132 of this issue, Kiyozumi *et al.* (*2*) identify and characterize the first known lumicrine factor, the germ cell–secreted protein neural epidermal growth factor–like like 2 (NELL2). They demonstrate a key role for NELL2 in driving development of the IS of mice, culminating in the production of IS-secreted proteases that are indispensable for sperm processing in the epididymis and thus male fertility.

The epididymis is a convoluted ductal system of up to 1 m in length in mice and 6 m in humans. Testicular sperm must transit the length of the epididymis to gain the capacity for fertilization, including activation of the machinery that drives motility as well as egg recognition and binding. Given that spermatozoa are transcriptionally and translationally silent, the process of sperm maturation is largely attributed to activities of epididymal epithelial cells, which can include post-translational modification, augmentation, and processing of sperm proteins. Proteomic comparisons of sperm collected from the more proximal regions of the epididymis

[1]Priority Research Centre for Reproductive Science, Discipline of Biological Sciences, the University of Newcastle, Callaghan, NSW, Australia. [2]Center for Reproductive Biology, School of Molecular Biosciences, Washington State University, Pullman, WA, USA. Email: joatley@wsu.edu

Published by AAAS

(caput) with sperm from the distal region (cauda) reveal differential expression in more than 50 proteins (*3*) as well as 77 phosphoprotein changes (*4*).

Given the importance of the epididymis in sperm maturation, aberrations in epididymal function can be catastrophic for male fertility. Although endocrine factors such as androgens are integral for epididymal development during embryonic and early postnatal life, studies with mice indicate that lumicrine factors are critical for differentiation of the IS from postnatal day 15 (P15) to P19 (*5*). Indeed, when the efferent ductiles that connect testicular seminiferous tubules and epididymal tubules are ligated during development, thus preventing the flow of lumicrine factors into the epididymis, the epithelial cells of the IS do not undergo differentiation and default into an apoptotic pathway (*6*, *7*). Although testicular lumicrine factors likely stimulate IS differentiation and proliferation through activation of SFKs (SRC proto-oncogene family kinases), ERKs (extracellular signal–regulated kinases), and AMPKs (adenosine monophosphate–activated protein kinases) within the epithelial cells (*5*), the identity of these lumicrine factors has remained elusive.

In 1999, an orphan receptor tyrosine kinase, c-ros oncogene 1 (ROS1), was identified as the first putative lumicrine receptor that is expressed in the IS of the mouse epididymis (*1*). Although spermatogenesis was grossly normal, development of the IS in *Ros1*-deleted mice was impaired, and their sperm were unable to fertilize an egg. This led to the hypothesis that factors produced in the postnatal testis transit to the epididymis to stimulate differentiation and development of the IS and drive epithelial cells to produce molecules essential for sperm maturation. Unanswered questions center on which testicular cell types (germ cell and/or somatic cell) are responsible for synthesis of lumicrine factors, what these factors are, and how they stimulate the epithelial cells of the IS to differentiate and influence sperm maturation.

Kiyozumi *et al.* provide evidence that lumicrine factors that drive IS development are produced by testicular germ cells (which produce spermatozoa). They find that NELL2 binds directly to the ROS1 receptor, and inactivation of the *Nell2* gene phenocopies the *Ros1*-deleted mice. Specifically, *Nell2*-deleted mice have a poorly differentiated IS and produce sperm that cannot transit the uterotubal junction in the female reproductive tract or bind the glycoprotein "shell" (called the zona pellucida) that surrounds the oocyte, thus culminating in infertility.

Kiyozumi *et al.* highlight that impaired IS development attributed to lumicrine interruption becomes appreciable from 2 to 3 weeks of age, coinciding with the appearance of the first spermatocytes in the testis. Mining testicular single-cell RNA-sequencing databases revealed that *Nell2* expression is indeed restricted to spermatocytes in adult mice. To consolidate the idea that spermatocyte-produced factors influence IS development, the authors demonstrate a previously unappreciated immaturity of the IS in genetically sterile mice that lack advanced germ cells. Together, these data provide compelling evidence for a previously undetermined interaction between germ cells and the somatic cells that line the epididymis and regulate sperm maturation (see the figure).

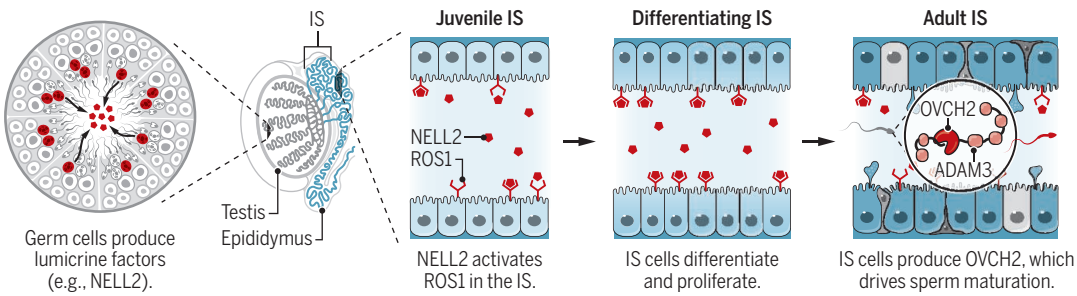
To understand how lumicrine factors prepare the IS epithelial cells to participate in sperm maturation, Kiyozumi *et al.* investigated the molecular events that follow NELL2-ROS1 binding. They identified the uncharacterized protease ovochymase 2 (OVCH2) that is secreted by the IS epithelial cells to cleave and thus activate a key protein involved in sperm-egg binding [A disintegrin and metallopeptidase domain 3 (ADAM3)]. Thus, as is the case with *Nell2*-, *Ros1*-, and *Adam3*-deleted mice, *Ovch2*-deleted mice produced morphologically normal sperm that are unable to transit the uterotubal junction in the female reproductive tract or bind the egg zona pellucida for fertilization.

The NELL2-ROS1-OVCH2-ADAM3 pathway characterized by Kiyozumi *et al.* is an eloquent example of lumicrine function that is likely to catalyze characterization of other similar molecular cross-talk networks that regulate male fertility. Certainly, it is possible that OVCH2 cleaves other key proteins in the spermatozoon that aid in its ascent to become fertilization competent. Further, it is likely that other lumicrine factors are at play to stimulate epididymal maturation and the secretion of alternate protein processing enzymes that drive sperm maturation during epididymal transit.

The characterization of lumicrine signaling between the germ cells and epididymis could profoundly alter the understanding of male infertility and, further, stimulate the development of new approaches for treating infertility or devising nonhormonal male contraceptives. In up to 80% of male infertility cases, sufficient numbers of sperm are produced, yet functionality is compromised (*8*, *9*). A major cause of such fertilization failures is an inability of the sperm to recognize and bind to the zona pellucida (*10*), similar to the phenotype of lumicrine-deficient mouse models. Given that the causes of up to 60% of male infertility cases are unknown (*8*), exploring defective "lumicrinology" will provide a new avenue for diagnosis and treatment. Moreover, targeting the lumicrine system may present an efficacious strategy for identifying druggable targets for the long sought-after male contraceptive pill. ■

Lumicrine regulation of epididymal development and function

Lumicrine factors, such as neural epidermal growth factor–like like 2 (NELL2), are produced by germ cells in the testis during maturation of the epithelial cells that line the initial segment (IS) of the epididymis through activation of c-ros oncogene 1 (ROS1). Mature IS cells produce ovochymase 2 (OVCH2), which cleaves a disintegrin and metallopeptidase domain 3 (ADAM3) on immature sperm to attain fertilization competency.



REFERENCES AND NOTES

1. C.-H. Yeung, E. Sonnenberg-Riethmacher, T. G. Cooper, *Biol. Reprod.* **61**, 1062 (1999).
2. D. Kiyozumi *et al.*, *Science* **368**, 1132 (2020).
3. M. A. Baker, L. Hetherington, G. Reeves, J. Müller, R. J. Aitken, *Proteomics* **8**, 2312 (2008).
4. M. A. Baker *et al.*, *J. Proteome Res.* **10**, 1004 (2011).
5. B. Xu, A. M. Washington, B. T. Hinton, *Biol. Reprod.* **95**, 15 (2016).
6. B. Xu, L. Yang, R. J. Lye, B. T. Hinton, *Biol. Reprod.* **83**, 807 (2010).
7. T. T. Turner, D. S. Johnston, J. N. Finger, S. A. Jelinsky, *Biol. Reprod.* **77**, 165 (2007).
8. M. Punab *et al.*, *Hum. Reprod.* **32**, 18 (2017).
9. M. G. Hull *et al.*, *Br. Med. J. (Clin. Res. Ed.)* **291**, 1693 (1985).
10. K. A. Redgrove, R. J. Aitken, B. Nixon, in *Binding Protein*, K. Abdelmohsen, Ed. (IntechOpen, 2012), pp. 73–122.

10.1126/science.abc2732

GRAPHIC: A. KITTERMAN/*SCIENCE*

Published by AAAS

OCEANOGRAPHY

Deep-ocean seafloor islands of plastics

The processes controlling sediment transport also concentrate microplastics

By **David Mohrig**

One of the most striking images of ocean pollution are the patches or islands of floating plastic debris, concentrated in open-ocean gyres (*1*) and large enough to be seen from space. These concentrations of plastics on the ocean surface were first recognized in the early 1970s (*2*). Even with the impressive size of these patches, mass-balance estimates for ocean-borne plastics pointed toward a sink. That sink has recently been shown to be deposition on the deep seafloor (*3*). On page 1140 of this issue, Kane *et al.* document the occurrence of enriched zones or islands of plastic debris that accumulate on the seafloor of the deep ocean (*4*).

Very different styles of plastics transport are associated with sea-surface and seafloor pollution. These styles are well understood for sediment transport on Earth's surface, which in turn are linked to surface evolution (*5*). Kane *et al.* document that the accumulation of microplastics on the Mediterranean seafloor is not simply passive vertical settling through the water column as has been commonly assumed, but rather represents reworking by deep-sea currents. These concentrate microplastics into patches or islands at predictable locations if the microplastics are treated as sediment that can be eroded, transported, and deposited by a deep-ocean flow field. The same processes that control patterns of erosion and deposition on terrestrial landscapes and shallow-marine shelves also pertain to plastics in the deeper ocean. The authors add much needed confirmation that concepts and methods developed from easier-to-access environments can be applied with confidence to remote undersea inquiries (*5*).

Although the primary focus of Kane *et al.* is to demonstrate the fate and focusing of pollutants in the deep sea, the authors also identify a likelihood for colocated hotspots in microplastic concentration and deep-ocean biodiversity. The same currents driving the enrichment of microplastics are efficient conveyers of nutrients and dissolved oxygen to the seafloor. Positioning of the colocated hotspots is affected by the submarine topography that guides the deep-ocean currents. This linkage is analogous to those already identified between surface transport and food webs on terrestrial landscapes (*5, 6*). This observation opens an opportunity to connect the spatial structure of deep-water ecosystems and pollutants to the spatial structure of surrounding submarine landscapes. Understanding the connections between seafloor pollution by microplastics and deep-sea ecology requires a unified science of Earth's surface dynamics that remains one of the great integrating challenges of environmental studies.

Determining the fate of plastic debris transported from land to sea (*7*) contributes to quantifying mass fluxes defining Earth's source-to-sink system. Research on sediment-routing systems (*8*, *9*) has already identified many of the complications and pitfalls inherent to tracking particles across the environment. Sediment-flux signals can be phase-shifted, lagged, and buffered by the internal dynamics of the transport system. These same dynamics will confound analyses of routing signals for plastics across Earth's surface and into the deep-ocean sink. Kane *et al.* demonstrate that microplastics deposited in this deep-sea environment are still subject to later erosion, transport, and redeposition due to time and space variations in the near-bed velocity fields of thermohaline or contour currents. Understanding the ultimate fate of these microplastics requires high-resolution submarine topography or bathymetry because the deep-sea currents interact with this topography to produce the spatial changes in velocity that set patterns of sediment and microplastic erosion, transport, and accumulation. Unfortunately, high-resolution bathymetric data simply do not exist at the global scale. The most widely used seafloor dataset is the General Bathymetric Chart of the Oceans, available on a 15–arc sec world grid (~463-m resolution at the equator). It is still surprising that this worldwide ocean product exists at a resolution poorer than that of the 200-m digital elevation model available for the entire surface of Mars (*10*). Until ocean-basin scale models improve, our detailed understanding of solids transport at and near the seafloor will be restricted to local or regional studies connected with higher-resolution bathymetric models.

> **"In less than 80 years, plastic debris has been...distributed across Earth's surface."**

Kane *et al.* identify extremely high concentrations for microplastics in deep-sea sediment drift deposits (*11*). This style of deposit has been the focus of previous studies because relatively rapidly accumulating drift deposits constitute an excellent archive of proxy measures for past Earth states (*12*). This correlation of highest plastic concentrations with high-fidelity paleoenvironmental records suggests the potential for a Global Boundary Stratotype Section and Point (*13*), if the Anthropocene is formally recognized as a geologic epoch (*14*). The first occurrence of microplastic in core from drift deposits could serve as a "golden spike" recording the lower boundary or beginning of the proposed epoch (*14, 15*). In less than 80 years, plastic debris has been effectively distributed across Earth's surface. Kane *et al.* demonstrate that even in the deep sea, its motion and accumulation is dominated by the feedbacks between fluid flow, sediment transport, and topography that are an overarching hallmark of Earth's surface system. ■

Jackson School of Geosciences, The University of Texas at Austin, Austin, TX USA. Email: mohrig@jsg.utexas.edu

REFERENCES AND NOTES

1. A. Cózar *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 10239 (2014).
2. E. J. Carpenter, K. Smith, *Science* **175**, 1240 (1972).
3. L. C. Woodall *et al.*, *R. Soc. Open Sci.* **1**, 140317 (2014).
4. I. A. Kane *et al.*, *Science* **368**, 1140 (2020).
5. C. Paola *et al.*, *Water Resour. Res.* **42**, W03S10 (2006).
6. M. E. Power, W. E. Dietrich, *Ecol. Res.* **17**, 451 (2002).
7. J. R. Jambeck *et al.*, *Science* **347**, 768 (2015).
8. P. A. Allen, *Nature* **451**, 274 (2008).
9. B. W. Romans, S. Castelltort, J. A. Covault, A. Fildani, J. P. Walsh, *Earth Sci. Rev.* **153**, 7 (2016).
10. R. L. Fergason *et al.*, "HRSC and MOLA Blended Digital Elevation Model at 200m v2, Astrogeology PDS Annex" (U.S. Geological Survey, 2018); http://bit.ly/HRSC_MOLA_Blend_v0
11. M. Rebesco, F. J. Hernández-Molina, D. Van Rooij, A. Wåhlin, *Mar. Geol.* **352**, 111 (2014).
12. A. C. Englebrecht, J. P. Sachs, *Geochim. Cosmochim. Acta* **69**, 4253 (2005).
13. J. Remane *et al.*, *Episodes* **19**, 77 (1996).
14. S. L. Lewis, M. A. Maslin, *Nature* **519**, 171 (2015).
15. J. Zalasiewicz *et al.*, *Anthropocene* **13**, 4 (2016).

ACKNOWLEDGMENTS

Thanks to N. Tull, P. Passalacqua, J. Hariharan, K. Wilson, H. Hassenruck-Gudipati, K. Wright, T. Jarriel, E. Prokocki, J. Swartz, and S. Rahman for brainstorming on plastics transport.

10.1126/science.abc1510

Published by AAAS

CHEMICAL PHYSICS

The link between electrolytes and metals

Photoelectron spectroscopy maps a gradual transition for solvated electrons in ammonia

By **Christine M. Isborn**

Solutions of alkali metals in liquid ammonia have fascinated scientists for more than 200 years. The "fine blue colour" of a dilute solution indicative of solvated electrons was first noted by Sir Humphry Davy in 1808 [(*1*), p. 63] and independently published by W. Weyl in 1864 (*2*). A bronze sheen as the solution becomes highly concentrated (see the figure) develops as these solvated electrons coalesce into a metallic continuum. Charles Kraus, an early pioneer of the study of these solutions (*3*), noted that "these solutions…constitute a link between electrolytes, on the one hand, and metals, on the other" [(*4*), p. 83]. On page 1086 of this issue, Buttersack *et al.* (*5*) combine low-temperature x-ray photoelectron spectroscopy with high-level simulations to reveal the energetics of metal-ammonia solutions across a large concentration range. These studies provide the missing energetic link to characterize the journey of solvated electrons from electrolyte to metal.

Electrons are usually either relatively localized in atomic or molecular orbitals or can be delocalized in the energy bands of extended solids. For the much less common case of solvated electrons—which exist in a number of solvents, including water, but are most stable in ammonia—questions remain about the extent of localization and the degree of association with the parent ions and surrounding solvent. The solvent environment may change to support these species, and as the concentration increases, the electrons may interact to become spin-paired dielectrons and ultimately coalesce into a metallic state.

These questions have motivated many studies of hydrated and ammoniated electrons over the years (*6*, *7*). One of the most effective techniques for studying electronic energy levels is photoelectron spectroscopy, which measures the kinetic energy of electrons emitted after a substance is irradiated with light (based on the principles of the photoelectric effect), revealing electronic binding energies. This technique has been used to characterize hydrated electrons and ammoniated electrons, albeit at low concentrations, well below the electrolyte-to-metal transition.

The development of a new apparatus enabled the application of x-ray photoelectron spectroscopy to a liquid ammonia microjet (*8*), allowing Buttersack *et al.* to collect the photoelectrons from the volatile, polar refrigerated metal-ammonia solution across the wide concentration range of 0.012 to 9.7 mol % metal. Over this range, the solution color changes from a lighter to a deeper blue and finally develops a bronze metallic sheen. The photoelectron spectra reveal the onset and growth of a peak beginning at a concentration of 0.08 mol % metal (see the figure). The energy of this peak is independent of the identity of the alkali metal in the solution, which suggests that the peak arises only from the ammoniated electron and that the metal parent ion does not play a direct role in this transition.

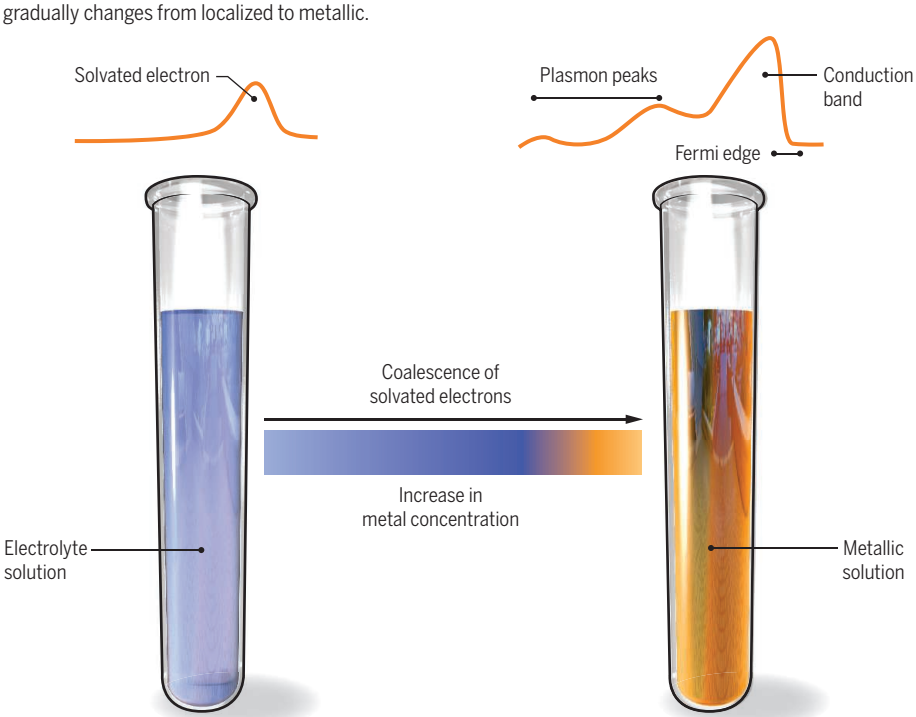
As the concentration of the alkali metal increases, this peak gradually grows into a metallic conduction band with a sharp Fermi edge, and an additional plasmon peak appears. This plasmon peak is responsible for the characteristic bronze color of the metallic solution. This gradual transition is in contrast to the sharp transition proposed in recent work on metal-ammonia nanodroplets (*9*), which shows that the bulk liquid and small solvent clusters support different solvated electron structures at higher concentrations and therefore different metallic onsets.

Modeling by Buttersack *et al.* complements the photoelectron spectroscopy data. They apply a metallic free-electron gas model to fit the growth of the conduction band and the sharp Fermi edge in the photoelectron spectra as the solution undergoes the electrolyte-to-metal transition. These

Chemistry and Chemical Biology, University of California Merced, Merced, CA 95343, USA.
Email: cisborn@ucmerced.edu

A fascinating color change

Alkali metal-ammonia solutions change from blue to bronze as the concentration of the metal is increased. This color change marks the shift of the solution from electrolyte to metallic. The photoelectron spectroscopy experiments by Buttersack *et al*. across the concentration range show how the electron gradually changes from localized to metallic.



Solvated electrons
A localized peak arises from the solvated electron. A similar feature appears for the dielectron.

Metallic electrons
At higher concentrations, metallic features appear in the spectrum at concentrations well below the visual color change to bronze.

GRAPHIC: C. BICKEL/*SCIENCE*

Published by AAAS

metallic spectral features begin to reveal themselves at concentrations lower than when the solution begins to visually appear metallic bronze.

On the dilute side of the concentration range (only one solvated electron or one solvated dielectron), ab initio molecular dynamics provide structures of the electron and dielectron solvated by ammonia molecules. These structures are then used for high-level vertical dissociation energy computations. Performing ab initio molecular dynamics of an excess electron or dielectron in bulk ammonia is a substantial computational feat and advances the field beyond previous static cluster calculations to reveal a diffuse ammonia solvation shell that is similar for both the electron and the dielectron. The spin densities suggest that the ammoniated electron resides within a cavity that is less structured than that of the hydrated electron. The high-level vertical dissociation energy calculations show that the energies to ionize the electron and the dielectron both fall within the measured photoelectron signal.

With greatly increased computational resources, ab initio molecular dynamics simulations could be expanded to a larger scale to include metal ions dissolved in ammonia at a variety of concentrations. Studies such as these may be necessary to resolve some of the remaining controversy of the localized versus delocalized nature of the hydrated and ammoniated electron (*10*–*13*). These results would provide additional atomistic and electronic details of the electrolyte-to-metal transition. For now, the spectroscopic studies by Buttersack *et al.* provide the missing energetic and spectroscopic link and reveal the gradual transition to metallic behavior before our eyes can see it. ■

REFERENCES AND NOTES

1. S. J. M. Thomas, P. P. Edwards, V. L. Kuznetsov, *ChemPhysChem* **9**, 59 (2008).
2. W. Weyl, *Ann. Phys.* **197**, 601 (1864).
3. C. A. Kraus, *J. Am. Chem. Soc.* **43**, 749 (1921).
4. C. A. Kraus, *J. Chem. Educ.* **30**, 83 (1953).
5. T. Buttersack *et al.*, *Science* **368**, 1086 (2020).
6. M. T. J. H. Lodge *et al.*, *J. Phys. Chem. B* **117**, 13322 (2013).
7. E. Zurek, P. P. Edwards, R. Hoffmann, *Angew. Chem. Int. Ed.* **48**, 8198 (2009).
8. T. Buttersack *et al.*, *J. Am. Chem. Soc.* **141**, 1838 (2019).
9. S. Hartweg, A. H. C. West, B. L. Yoder, R. Signorell, *Angew. Chem. Int. Ed.* **55**, 12347 (2016).
10. I. A. Shkrob, *J. Phys. Chem. A* **110**, 3967 (2006).
11. R. E. Larsen, W. J. Glover, B. J. Schwartz, *Science* **329**, 65 (2010).
12. L. D. Jacobson, J. M. Herbert, *Science* **331**, 1387 (2011).
13. C.-C. Zho, E. P. Farr, W. J. Glover, B. J. Schwartz, *J. Chem. Phys.* **147**, 074503 (2017).

ACKNOWLEDGMENTS

The author is supported by grants from the U.S. Department of Energy, Basic Energy Sciences, Computational and Theoretical Chemistry and Condensed Phase and Interfacial Molecular Science programs (DE-SC0019053 and DE-SC0020203).

10.1126/science.abb9717

NEUROSCIENCE

Unblinding with infrared nanosensors

Gene therapy and nanotechnology come together to fight degenerative blindness

By **Katrin Franke** *and* **Anna Vlasits**

Many cases of blindness result from progressive loss of photoreceptors, which are the light-sensing cells in the eye. For individuals with such progressive blindness, potential therapies aim at restoring vision by making the retina light-sensitive again while minimally interfering with any healthy photoreceptors—goals that are usually contradictory. Many current therapeutic strategies interfere with remaining vision, making them primarily suitable for patients who have lost all light sensitivity. On page 1108 of this issue, Nelidova *et al.* (*1*) present a potential solution to this conundrum: making the retina sensitive to infrared light, which is largely undetectable by human photoreceptors. They use engineered nanoparticle sensors and gene therapy to induce infrared light sensitivity in mice with inherited degenerative blindness and in postmortem human retinas. This approach might avoid damage to functional photoreceptors by preventing saturation or hyperactivation while inducing light sensitivity in patients with partial retinal degeneration.

At the level of the retina, a multilayered array of more than 100 types of neurons sorts complex visual features, such as motion and color, into separate channels to send to the brain (*2*). When photoreceptors fail, the entire downstream network is affected, and restoring the visual system's physiological function becomes challenging. In addition, mammalian rod and cone photoreceptors, unlike those of some species in the animal kingdom, cannot regenerate. For patients with degenerative blindness, the aim of therapy is therefore twofold: to slow degeneration while preserving remaining vision, and to restore some vision once degeneration is complete. Several types of therapies show promise in slowing degenerative blindness, including targeted gene therapies and stem cell–based approaches (*3*, *4*). One example is Leber congenital amaurosis, a disease for which a gene therapy has led to an improvement in patients' visual function (*5*). Other therapeutic strategies include electrical implants and optogenetic gene therapies, which aim at restoring vision (*4*, *6*).

Institute for Ophthalmic Research, Bernstein Center for Computational Neuroscience, Center for Integrative Neuroscience, Tübingen University, 72076 Tübingen, Germany. Email: katrin.franke@uni-tuebingen.de

The optogenetic approach induces the expression of light-sensitive ion channels through gene therapy to restore light sensitivity to retinal neurons. Several molecular candidates can restore light sensitivity in animals and in postmortem human retinas (*7*), and some are now in clinical trials. But currently these molecules require much more light than normal photoreceptors to become activated, and most therapies would require video goggles to boost the effective brightness of incoming images. For patients with some remaining vision, this strategy would saturate or possibly even damage their remaining functional photoreceptors.

Nelidova *et al.* circumvent these problems by making the retina sensitive to infrared light, which is light beyond the visible spectrum and emitted by, for example, warm objects. Their approach combines gene therapy with the use of gold nanorods, an emerging nanotechnology for activating molecules in the human body (*8*). Nelidova *et al.* use gold nanorods as antennae for infrared light, transforming the light into heat through a process called surface plasmon resonance. Genetic constructs injected into the eye then cause the expression of temperature-sensitive transient receptor potential (TRP) channels in photoreceptors. Such TRP channels are normally found in mammalian heat-sensing nerves in the skin, as well as in the infrared-sensing organs of some snakes and vampire bats, and are able to transform heat into electrical changes in the membranes of cells (*9*). The authors use antibodies to link the heat-emitting gold nanorods to heat-sensitive TRP channels. Thus, infrared light can activate photoreceptors (see the figure).

To test whether this strategy can restore visual function, Nelidova *et al.* express TRP channels in cone photoreceptors of a mouse model of degenerative blindness. They find that neural activity measured in the retina and visual cortex correlated with infrared light stimuli. In addition, they show that treated blind mice can use their infrared light sensitivity to learn a simple visually guided

Published by AAAS

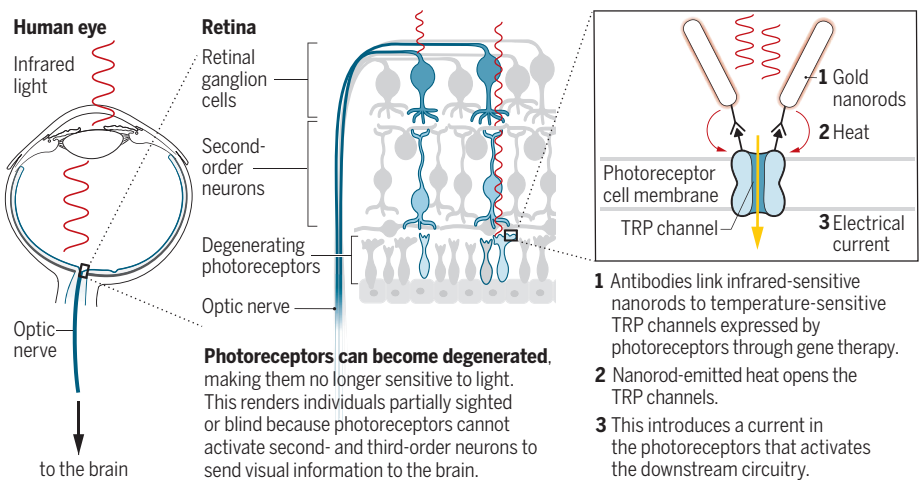
behavior. The authors tested their nanorod–TRP channel approach in cultured, light-insensitive postmortem human retinas, demonstrating that it introduces infrared light sensitivity to this tissue—a critical step in evaluating its relevance for human patients.

A different nanotechnology, called up-conversion nanoparticles, which binds to photoreceptors and "up-converts" infrared into visible light, can make photoreceptors virtually infrared-sensitive (*10*). Although degenerating photoreceptors would likely not be able to use up-conversion nanoparticles, mice treated with this technology used their infrared sensitivity to perform complex visual tasks including shape recognition. Such detailed behavioral evaluation is critical, because it is not clear to what extent the already-developed brain can interpret a new sensory modality to guide behavior—although studies support some plasticity of the adult mammalian brain for integrating new sensory input (*11*, *12*). Both examples demonstrate the strengths of nanotechnology tools over other methods. These tools are more light-sensitive than conventional optogenetics, approaching the sensitivity needed to work under normal daylight levels. Because the nanoparticles harness a different wavelength of light, it might be possible for normal vision and infrared vision to operate in parallel.

The nanorod–TRP channel approach used by Nelidova *et al.* faces further challenges before it can reach the clinic. It is promising that gold nanorods have, so far, appeared to be safe in humans (*13*). Similarly, ocular gene therapies seem to be low-risk and effective (*3*, *4*). However, the main challenge of any ocular gene therapy is to improve the efficiency and completeness of gene introduction (*3*) as well as to provide long-term effectiveness (*14*). More specifically, because nanorods and TRP channels cannot currently be targeted selectively to degenerating photoreceptors, the interaction between induced infrared sensitivity and the intrinsic light sensitivity of healthy photoreceptors requires further investigation. In addition, many objects that humans see are not necessarily infrared-emitting or infrared-reflecting; thus, goggles to convert visible light to infrared light would likely be necessary. Nonetheless, this system has exceptional promise for basic research. Tools that can reintroduce light sensitivity to postmortem human retinas (*15*) offer the potential for studying human retinal function in much greater detail than was previously possible. This basic knowledge is important for any approach to vision restoration, as it would reveal what kinds of functions need to be restored. ■

Nanorods and heat-sensing proteins for infrared detection

Injecting the eye with infrared-sensitive nanorods and genetic constructs to induce the expression of temperature-sensitive transient receptor potential (TRP) channels in photoreceptors may confer infrared vision.



ACKNOWLEDGMENTS

The authors are supported by the German Research Foundation (EU 42/9–1), the German Federal Ministry of Education and Research (01GQ1002), and the Max Planck Society (M.FE.A.KYBE0004).

10.1126/science.abc2294

REFERENCES AND NOTES

1. D. Nelidova *et al.*, *Science* **368**, 1108 (2020).
2. U. Grünert, P. R. Martin, *Prog. Retin. Eye Res.* 10.1016/j.preteyeres.2020.100844 (2020).
3. D. Dalkara *et al.*, *Hum. Gene Ther.* **27**, 134 (2016).
4. B. Roska, J.-A. Sahel, *Nature* **557**, 359 (2018).
5. A. Maguire *et al.*, *N. Engl. J. Med.* **358**, 2240 (2008).
6. E. Zrenner, *Science* **295**, 1022 (2002).
7. M. P. Simunovic *et al.*, *Exp. Eye Res.* **178**, 15 (2019).
8. Z. Qin, J. C. Bischof, *Chem. Soc. Rev.* **41**, 1191 (2012).
9. L. J. Hoffstaetter *et al.*, *Pflugers Arch.* **470**, 745 (2018).
10. Y. Ma *et al.*, *Cell* **177**, 243 (2019).
11. E. E. Thomson *et al.*, *Nat. Commun.* **4**, 1482 (2013).
12. K. Mancuso *et al.*, *Nature* **461**, 784 (2009).
13. F. Masse *et al.*, *Med. Res. Rev.* **39**, 302 (2019).
14. S. G. Jacobson *et al.*, *N. Engl. J. Med.* **372**, 1920 (2015).
15. A. Sengupta *et al.*, *EMBO Mol. Med.* **8**, 1248 (2016)

SYNTHETIC BIOLOGY

Follow the barcoded microbes

Genetically engineered spores give object provenance technology new avenues

By **Jeff Nivala**

Tracking where a physical object originated and where it has been, known as object provenance, is becoming increasingly important with the globalization of supply chains (*1*). Object-labeling technologies that are scalable, robust, and difficult to falsify would support, for example, the mitigation of foodborne illness outbreaks and the manufacture of counterfeit goods. However, there currently exists no single tagging method that satisfies all these requirements. On page 1135 of this issue, Qian *et al.* (*2*) describe a new tagging technique, called the barcoded microbial spores (BMS) system, that uses genetically engineered microbes as molecular tags to address the object provenance problem. The authors demonstrate that BMS can label a range of surfaces and persist for months in real-world conditions. Furthermore, they show how this technology can tag objects that come into only brief contact with BMS-labeled surfaces, suggesting the utility of BMS in forensic surveillance applications (*3*).

Microbes are ubiquitous in our everyday environments, with different geographical locations having distinct microbial populations. Physical objects can even take on the microbial signature of their environments over time (*4*, *5*), which has led to the suggestion that this naturally occurring signature could be used for object provenance (*6*). Although it would be convenient to have environmental microbes acting as automatic label-makers, actually implementing such an approach would pose a number of challenges. For instance, extensive environmental mapping of microbes would have to be done first. This would be expensive and

Molecular Information Systems Lab, Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, USA. Email: jmdn@uw.edu

GRAPHIC: N. CARY/*SCIENCE*

Published by AAAS

time-consuming. Furthermore, environmental microbial abundances can be similar at disparate locales and also change over time, making the use of natural microbe labels a nonstarter for many applications.

Instead, Qian *et al*. take this concept into the realm of synthetic biology by engineering synthetic strains of microbes to function as molecular labels. To do this, they inserted short, specific DNA sequences (barcodes) into the genomes of the bacterium *Bacillus subtilis* and the yeast *Saccharomyces cerevisiae*. They designed and experimentally validated dozens of barcode sequences that could be used in combinations to generate an essentially infinite number of potential tag sets. To tag an object, a set of barcoded strains were mixed together in spore form and sprayed on the surface of an object. A spore is a physically tough, inactive cell state that can persist in harsh environmental conditions without growth (*7*). This resilience allows the barcoded microbial spores to endure in diverse ecosystems without rupturing and exposing their DNA barcodes to the elements, risking tag loss.

To identify the barcodes, the surface of the object is swabbed to collect a sample, which is prepared for input to one of a number of different decoding devices, including a DNA sequencer or a quantitative polymerase chain reaction machine. In addition to these readout options, Qian *et al*. also focused on using a nascent DNA detection technology called SHERLOCK (*8*), which is more amenable to deployment in field settings. This feature is critical to many potential provenance applications. BMS decoding should also be compatible with other field-deployable DNA-sensing technologies, such as commercial nanopore devices (*9*), although the authors did not explore this compatibility.

The release of genetically modified organisms into uncontrolled environments may come with risks. So, Qian *et al*. built clever safeguards into the BMS system to prevent the unintended spread and proliferation of their microbial tags in the environment, beginning with careful selection of the microbial species themselves. Both *B. subtilis* and *S. cerevisiae* are commonly found in the environment and in food samples, and products derived from them have already been granted "generally recognized as safe" (GRAS) status by the U.S. Food and Drug Administration. To prevent their proliferation in native environments, Qian *et al*. used auxotrophic *B. subtilis* and *S. cerevisiae* strains that require supplementation of key amino acids for growth. The authors took additional steps to prevent the microbial spores from returning to an active cellular state by either removing genes essential to this process (for *B. subtilis*) or by boiling the spores to permanently heat-inactivate them (*S. cerevisiae*). These measures were adequate to ensure that the spores did not replicate even in the most favorable laboratory conditions. As there is still the potential for auxotrophs to grow by scavenging for metabolites in the environment, these additional safeguards are prudent. An alternative strategy to biocontainment would be the use of so-called "recoded" cell strains that have been engineered to be dependent on completely synthetic amino acids that are not found in natural ecosystems (*10*, *11*).

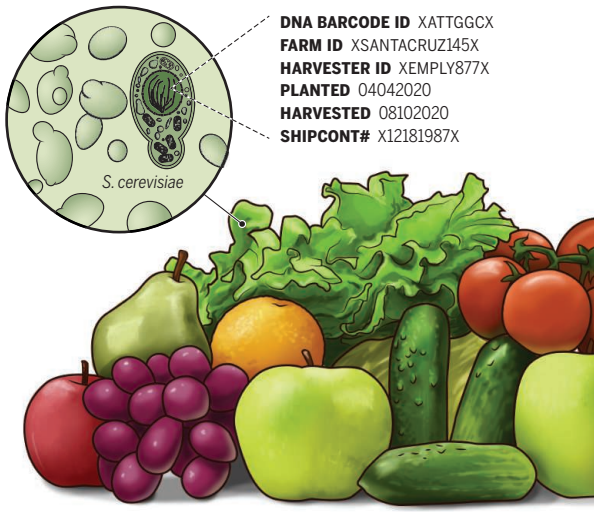
The value of a provenance system depends on its lifetime and applicability to different objects. The BMS system was prototyped on an impressive number of surfaces and simulated environments, including sand, soil, carpet, and wood, in addition to an outdoor grass area. Notably, the barcoded microbes persisted and remained detectable on these surfaces for months, even after real or simulated weather conditions and physical perturbations such as vacuuming and sweeping. Beyond typical object provenance, in which the labels are applied directly to the object being tracked, Qian *et al*. also demonstrated that BMS tags could be transferred to other objects that come in transient contact with a labeled surface. The authors found that they could even trace the path of shoes that had walked on a BMS-tagged floor. The potential use for such a technology by, for example, law enforcement is compelling, although additional work is needed to determine how the BMS system withstands more rigorous real-world stresses, such as the application of cleaning agents to the surface.

Perhaps the most powerful use of microbial tags will come from their application to agricultural and other food supply chains. For example, the tags could safely be sprayed directly on food products, as Qian *et al*. showed with leafy plants (see the figure). *Bacillus thuringiensis*, which is closely related to *B. subtilis*, is already in common use as an insecticide for agricultural products that are commercially available today (*12*). The authors found that *B. thuringiensis* genomic DNA can be detected on store-bought produce even after washing, boiling, frying, and microwaving, thus highlighting the hardiness of DNA-based tags. As would be expected, other microbe and DNA-based provenance systems are also being developed, in both academia (*13*) and industry. We may well be on a path to a brave new world in which food provenance tracking goes not just from the farm to the table, but all the way to the sewer (*14*). ■

"Perhaps the most powerful use of microbial tags will come from their application to agricultural and other food supply chains."

Produce provenance data encoded in the DNA of a synthetic microbe

Genetically barcoded spores could be used for a new generation of object (e.g., food) provenance applications.



GRAPHIC: V. ALTOUNIAN/*SCIENCE*

REFERENCES AND NOTES

1. P. M. Wognum *et al.*, *Adv. Eng. Inform.* **25**, 65 (2011).
2. J. Qian *et al.*, *Science* **368**, 1135 (2020).
3. J. Gooch, B. Daniel, V. Abbate, N. Frascione, *Trends Analyt. Chem.* **83**, 49 (2016).
4. S. Lax *et al.*, *Science* **345**, 1048 (2014).
5. C. Jiang *et al.*, *Cell* **175**, 277 (2018).
6. S. Lax *et al.*, *Microbiome* **3**, 21 (2015).
7. N. Ulrich *et al.*, *PLOS ONE* **13**, e0208425 (2018).
8. J. S. Gootenberg *et al.*, *Science* **356**, 438 (2017).
9. M. Jain, H. E. Olsen, B. Paten, M. Akeson, *Genome Biol.* **17**, 239 (2016).
10. D. Mandell *et al.*, *Nature* **518**, 55 (2015).
11. A. J. Rovner *et al.*, *Nature* **518**, 89 (2015).
12. G. Sanahuja *et al.*, *Plant Biotechnol. J.* **9**, 283 (2011).
13. K. Doroschak *et al.*, bioRxiv 10.1101/2020.03.06.981514 (2020).
14. R. J. Newton *et al.*, *mBio* **6**, e02574 (2015).

10.1126/science.abc4246

Published by AAAS

VIEWPOINT: COVID-19

Serology assays to manage COVID-19

Measurement of antibodies to SARS-CoV-2 will improve disease management if used correctly

By **Florian Krammer**[1] *and* **Viviana Simon**[1,2,3]

In late 2019, China reported a cluster of atypical pneumonia cases of unknown etiology in Wuhan. The causative agent was identified as a new betacoronavirus, called severe acute respiratory syndrome–coronavirus 2 (SARS-CoV-2), that causes coronavirus disease 2019 (COVID-19) (*1*). The virus rapidly spread across the globe and caused a pandemic. Sequencing of the viral genome allowed for the development of nucleic acid–based tests that have since been widely used for the diagnosis of acute (current) SARS-CoV-2 infections (*2*). Development of serological assays, which measure the antibody responses induced by SARS-CoV-2 infection (past but not current infections), took longer. This is in part due to bottlenecks with availability of positive control sera and the need for extensive specificity and sensitivity testing in the context of preexisting immunity to seasonal coronaviruses. Serological assays are important for understanding the prevalence of and immunity to SARS-CoV-2.

Many types of serological assays have been developed over the past decades to measure antibody responses to pathogens in bodily fluids, especially blood serum or plasma. These assays use different platforms, including binding assays such as enzyme-linked immunosorbent assays (ELISAs), lateral flow assays, or Western blot–based assays. In addition, functional assays that test for virus neutralization, enzyme inhibition, or bactericidal assays can also inform on antibody-mediated immune responses. Collectively, serological assays are essential tools in the management of infectious diseases, including diagnosis of infection, measurements of protective antibody titers upon vaccination, and seroprevalence assessments of immunity in a population.

Serological assays for SARS-CoV-2 are now becoming widely available and include ELISAs (*3*–*7*), lateral flow assays (*5*, *8*, *9*) (see the figure), and virus neutralization assays. ELISA and lateral flow assays are performed with recombinant antigens, such as the spike protein (the main surface glycoprotein that is used to attach and enter cells) of SARS-CoV-2; the receptor-binding domain (RBD), which is part of the spike protein; or the viral nucleoprotein. Of note, using the SARS-CoV-2 nucleoprotein is expected to induce more cross-reactivity (antibodies that bind to multiple strains of coronavirus) than the spike protein, owing to sequence homology of the viral nucleoprotein. These assays can be handled at biosafety level 2 (and therefore can be carried out more widely), given the recombinant nature of the selected antigens. By contrast, neutralization assays with replication-competent SARS-CoV-2 have to be performed in biosafety level 3 facilities, which limits their application. Safer and more high-throughput alternatives to using infectious virus are under development and include the use of pseudotyped viral particle assays, in which the SARS-CoV-2 spike protein is grafted onto harmless viruses or virus-like particles.

A limited number of ELISA and lateral flow assays have recently received emergency use authorization from the U.S. Food and Drug Administration (FDA). In addition, many lateral flow assays from different companies are available, but their usefulness is questionable, given the lack of official performance validation with respect to sensitivity (how many true positives are detected) and specificity (the proportion of false positives) (*9*–*11*). Using serological assays with validated sensitivity and specificity performance is critical for obtaining meaningful results. For some applications, such as serosurveys in high-prevalence populations, somewhat lower specificity is acceptable, whereas sensitivity should be high. For uses where a false-positive test result would be consequential, very high specificity is essential. In general, both sensitivity and specificity should be as high as possible.

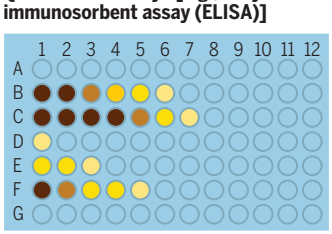
An important application of serological tests is to understand the antibody responses mounted upon SARS-CoV-2 infection and vaccination. Assays that inform on antibody titer and/or show antibody functionality (e.g., virus neutralization) will be extremely useful to answer important scientific questions about immune protection from reinfection. For example, do all infected individuals mount a robust antibody response to SARS-CoV-2 infection? It is unclear whether there is a difference in the antibody responses found in individuals presenting with severe, mild, and asymptomatic COVID-19 and how long antibody responses last. Moreover, it is unknown if the presence of binding antibody to the spike or RBD antigens correlates with virus neutralization. Whether antibody titers (binding or neutralizing) correlate with protection from reinfection is also unclear. Such data will be important when dissecting antibody responses generated by natural infection compared to vaccination.

Serological testing can also inform on the

Quantitative and binary readouts in serology assays

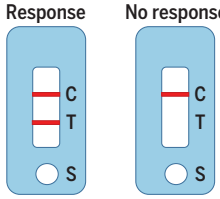
Quantitative and binary serology tests can provide important information about infection.

Quantitative assays [e.g., enzyme-linked immunosorbent assay (ELISA)]



Titers	Protected?
Negative	No
Titer 1:12,150	Yes
Titer 1:36,450	Yes
Titer 1:50	No
Titer 1:450	No
Titer 1:4050	Yes
Negative	No

Assay with binary result (e.g., lateral flow assay)



	Quantitative assays [e.g., enzyme-linked immunosorbent assay (ELISA)]	Assay with binary result (e.g., lateral flow assay)
Result	Quantitative titer	Yes or no
Linked to protection?	A quantitative titer can be linked to protection	A positive result can be loosely associated with protection
Could predict protection duration?	Yes	No
Scalability	Moderate	High
Ease of use	Performed in specialized laboratories	Easy to use, even as point-of-care test

GRAPHIC: H. BISHOP/*SCIENCE*

[1]Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [2]Division of Infectious Diseases, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [3]Global Health and Emerging Pathogens Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. Email: florian.krammer@mssm.edu; viviana.simon@mssm.edu

Published by AAAS

prevalence of SARS-CoV-2 infections in different populations. Although it is impractical to test the whole population, well-designed serosurveys are essential to determine how prevalent COVID-19 is in the general population, in selected subsections of the population (e.g., health care workers), or in specific risk groups. Both quantitative assays and assays with a binary outcome can be used for these surveys. Quantitative assays may provide more reliable results [e.g., two-step ELISAs (*12*)], but they are also harder to scale because they often have to be performed in specialized laboratories. By contrast, assays with binary outcomes (e.g., lateral flow assays) can be easily scaled and implemented because they are often point-of-care tests. Analyses of the results of serosurveys need to account for the sensitivity and specificity of the assay used as well as the estimated prevalence of infections in a population. In addition, biological variables resulting from in-depth characterization of the immune responses such as, but not limited to, the duration of the immune responses and the dynamic nature of antibody titers linked to severe, mild, and asymptomatic COVID-19 manifestations will need to be factored into calculating prevalence based on serosurveys. Currently, many of these critical variables are unknown, and any serosurvey analysis generated in the immediate future should be interpreted with caution.

Donors for convalescent plasma therapy can be identified with serology testing. Antibody-rich plasma or serum from convalescent individuals (or animals) has been used to treat many infections as well as snake bites. One of the earliest examples is the treatment of diphtheria with antiserum obtained from horses for which Emil von Behring received the Nobel Prize in 1901. More recently, antiserum has been used for the treatment of a range of viral infections (e.g., infections with Hantaan virus, Junin virus, measles virus, and Ebola virus, as well as potential rabies infections). Individuals who recover from COVID-19 develop antibodies to SARS-CoV-2. During the initial stages of the COVID-19 epidemic in China, convalescent plasma therapy was used compassionately (*13*) and has since been implemented in the United States and elsewhere. The success of this intervention likely increases with the antibody titer of the donor. It is, therefore, important to screen potential convalescent donors so that individuals with the highest antibody titers can be selected. This screening can be accomplished by measuring virus-neutralizing activity of the plasma, which is a lengthy process (several days) and needs to be performed in a biosafety level 3 laboratory. ELISA-based antibody testing that produces a titer is quick (hours) and easy to perform. Quantitative measurements of antibody titers from at least two different ELISAs have been shown to correlate well with neutralizing titers (*3, 4*).

Identifying individuals who are immune is an important but also complex and politically charged application of serological assays. Individuals who were infected with "common cold" human coronaviruses develop antibody responses and are protected from reinfection for a certain period of time, likely for years (*14*). If reinfection occurs, it is often mild or asymptomatic. In addition, infection with SARS-CoV-1 was shown to induce neutralizing antibody responses that last for several years (*14*). On the basis of these data, individuals with antibodies to SARS-CoV-2 are assumed to be less susceptible to reinfection, reducing the risk of severe COVID-19 and also limiting the possibility of spreading the virus. Therefore, it has been proposed that individuals with robust antibody responses could safely return to normal life and work, slowly starting the economy on a path to recovery. Detection of protective immune responses is also an important consideration for health care workers. In addition, people immune to SARS-CoV-2 could be spared from quarantine and social distancing measures during a potential second or third wave of SARS-CoV-2 infections in the winter of 2020. Accordingly, some countries have proposed an "immune passport" for such individuals.

However, there are numerous caveats that should be carefully considered before proceeding. It needs to be demonstrated that individuals who have developed antibodies to SARS-CoV-2 are protected. If antibodies provide immunity and protection, it is not (yet) known how long they will persist at the needed titer. A person protected today might no longer be protected in 6 months. It is, therefore, a matter of urgency to conduct studies aimed at dissecting the magnitude, duration, and functionality of the immune responses induced by SARS-CoV-2 infection, including antibodies, as well as cellular (adaptive) immune responses, and to determine the correlation between immune response and protection. In the absence of this knowledge, decisions about deploying the workforce may be based on incomplete information and guided by incorrect assumptions.

A known antibody titer that correlates with protection would also be extremely beneficial for vaccine development. Protective titers and/or correlates of immune protection have been established for many virus infections, including influenza virus, hepatitis A virus, hepatitis B virus, and measles virus. For several of these infections, the dynamics of the immune responses are well understood and the duration of protection based on antibody titers has been successfully modeled (*15*). For these types of studies, serological assays that measure a quantitative antibody titer have been instrumental. However, when converting the concept of an "immune passport" to practice, point-of-care serological assays that produce a binary response may also be useful. A combined strategic approach may be the safest while also being feasible. To account for sensitivity and false positives, if every positive lateral flow test result is confirmed with a second test that produces a titer—which also indicates the robustness of the response and could be linked to the presence and duration of protection—the number of false-positive results would be greatly reduced. Such a targeted sequential approach would provide reliable information on immunity and avoid putting individuals at risk.

Several academic laboratories have developed robust, specific serological assays, and high-quality commercial options are becoming available. In accordance with academic grassroots traditions, a toolkit to set up antibody assays has been distributed to more than 200 laboratories across the world, and a detailed protocol to facilitate local implementation has been published (*12*). With high-quality serological assays now available, the key challenge will be to apply and deploy these tests in a strategic manner to safely bring communities out of the current pandemic response back to the realm of "normal" life. ■

REFERENCES AND NOTES

1. P. Zhou *et al.*, *Nature* **579**, 270 (2020).
2. F. Wu *et al.*, *Nature* **579**, 265 (2020).
3. F. Amanat *et al.*, *Nat. Med.* 10.1038/s41591-020-0913-5 (2020).
4. N. M. A. Okba *et al.*, *Emerg Infect Dis* 10.3201/eid2607.200841 (2020).
5. B. Lou *et al.*, *medRxiv*, 10.1101/2020.03.23.20041707 (2020).
6. W. Tan *et al.*, *medRxiv*, 10.1101/2020.03.24.20042382 (2020).
7. J. Xiang *et al.*, *medRxiv*, 10.1101/2020.02.27.20028787 (2020).
8. Z. Li *et al.*, *J Med Virol*, 10.1002/jmv.25727 (2020).
9. R. Lassaunière *et al.*, *medRxiv*, 10.1101/2020.04.09.20056325 (2020).
10. J. D. Whitman *et al.*, *medRxiv*, 10.1101/2020.04.25.20074856 (2020).
11. E. R. Adams *et al.*, *medRxiv*, 10.1101/2020.04.15.20066407 (2020).
12. D. Stadlbauer *et al.*, *Curr. Protoc. Microbiol.* **57**, e100 (2020).
13. C. Shen *et al.*, *JAMA* **323**, 1582 (2020).
14. A. T. Huang *et al.*, *medRxiv*, 10.1101/2020.04.14.20065771 (2020).
15. N. Hens *et al.*, *Vaccine* **32**, 1507 (2014).

ACKNOWLEDGMENTS

The authors are supported by the National Institute of Allergy and Infectious Diseases (NIAID) Collaborative Influenza Vaccine Innovation Centers (CIVIC; 75N93019C00051), the NIAID Centers of Excellence for Influenza Research and Surveillance (CEIRS; HHSN272201400008C), institutional funding, and philanthropic donations. The SARS-CoV-2 reagents developed in our labs are freely available to the scientific community (www.beiresources.org/). Mount Sinai is in the process of commercial licensing of assays described here and has filed for patent protection.

Published online 15 May 2020
10.1126/science.abc1227

Published by AAAS

RETROSPECTIVE

Donald Kennedy (1931–2020)

Eminent public intellectual who served science and society

By **Gretchen C. Daily** and **Paul R. Ehrlich**

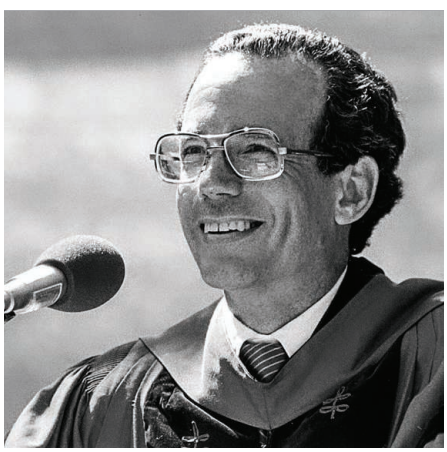
Donald Kennedy—public servant, university president, and former *Science* editor-in-chief—died on 21 April. He was 88. A born naturalist, broad intellectual, and leader, Don engaged in controversial areas as diverse as the safety of artificial sweeteners, the overuse of pesticides and antibiotics, the impacts of human population growth, the teaching of evolution, the philosophical basis for valuing nature, and academic duty. He served with brilliance, intense energy, and effectiveness. His death is a great loss to the United States and the world at this crucial moment in history.

Born in New York City in 1931, Don received all of his degrees (bachelor's, master's, and Ph.D.) in biology from Harvard University. He taught at Syracuse University for 4 years and then moved in 1960 to Stanford University, to which he devoted 32 years of his career. After leaving the laboratory bench, he took leadership roles at the university, in government, and finally at the journal *Science*.

In graduate school, Don explored the relation between brain and behavior. He devised a strategy for working on the simple nervous systems of invertebrate animals such as crayfish and lobsters—a choice that he liked to boast was superior to my (P.R.E.'s) choice of butterflies, in that Don could eat his specimens. He helped determine how restricted networks of neurons give rise to sensory perception and behavioral acts, such as swimming and flying. His presumption that the same rules of wiring neurons together apply to crayfish and humans has been justified time and again.

Don's research program advanced principles that later became widespread and part of modern systems neuroscience. His anthropocentric colleagues, mostly in medical schools, resisted his paradigm, complaining that scientists should be studying "real animals." His former student Ron Hoy reminisces that, in response, Don sardonically referred to his bottom-up methods based on invertebrates as the study of "virtual animals." He showed that stimulating just one or a few command neurons produces a complex sequence of movements. His laboratory was also instrumental in cell and network visualization techniques, using fluorescent dyes to define the functional connectivity within behavioral circuits.

In 1973, I (P.R.E.) and psychiatrist David Hamburg brought Don in to help organize Stanford's Program in Human Biology. Don soon took charge of the program and proceeded to use his uncommonly effective persuasive powers to convince me to teach for the program for almost a decade, despite my intention of staying only a few years. It was hard to say no to Don; he exuded genuine warmth, both with friends and upon first encounter, drawing people in with a light touch on the arm. Jeanne Kennedy, his first wife, recalls how "he could see the best in someone, who they really wanted to be…and make them feel as if they were that best version."



Don served as commissioner of the Food and Drug Administration from 1977 to 1979 before returning to Stanford as provost and then becoming its eighth president from 1980 to 1992. He led the university's Centennial Campaign, which brought in nearly $1.3 billion, the largest sum ever raised for higher education at the time. His enduring investments strengthened undergraduate teaching and financial aid, overseas study, and public service—including the launch of the Bing Stanford in Washington Program, an undergraduate internship in the U.S. capital. Later, Don weathered a challenging disagreement with the federal government over research reimbursement. Stanford was ultimately vindicated, but the publicity took a heavy toll.

In his professor days, Don was a legendary teacher and research mentor, inspiring large cohorts of undergraduates, graduate students, and postdocs, who themselves became leaders in science. He continued his mentoring as university president. I (G.C.D.) will never forget his addressing my freshman class, asking plainly, "Do you feel completely out of place?" and contending, "Every one of you should be uncomfortable! That's what'll make you grow!" He extended an open invitation to students to join him on his 6 a.m. runs and share what was on their minds. He would give students advice such as "Great plan, but you'll never make a three-point shot standing right under the basket."

Don's interests ranged from the practical to the esoteric, but he always looked at the big picture. He wrote on wide-ranging subjects, from nuclear war and pandemics to "what we don't know." Perhaps most unusually, in their book *Humans, Nature, and Birds*, he and coauthor Darryl Wheye explored the sweeping 30,000-year history of the depiction of birds in art.

During his tenure as editor-in-chief of *Science* from 2000 to 2008, Don greatly improved its coverage of ecology, evolution, and conservation at a time of rapidly intensifying human impacts on the biosphere. In doing so, he helped move problems such as land transformation, biodiversity loss, disruption of the nitrogen cycle, climate change, and ocean acidification to the top of the scientific agenda. He had contagious enthusiasm and boundless curiosity about every field the journal covered, but he was particularly excited when a paper touched on one of his personal passions. One day he called me (P.R.E.) and said, "We've just gotten a manuscript reporting ivory-billed woodpeckers in Arkansas." We started planning an expedition to see them, but the report of that now-extinct species turned out to be a false alarm, depriving us both of one final joint field adventure.

It is sadly ironic that coronavirus disease 2019 (COVID-19) killed Don. His grasp of the complex historical, demographic, environmental, epidemiological, economic, social, and political dimensions of pandemic risk was incomparable. He would have been on the front lines, advancing a coherent and effective response by the United States. Don's approach to every challenge was first and foremost based on science and evidence. He leavened this razor-sharp insight with a healthy sense of humor and humanity that earned him respect and admiration and made him beloved by those whose lives he touched. ■

Department of Biology, Stanford University, Stanford, CA 94305, USA. Email: gdaily@stanford.edu; pre@stanford.edu

10.1126/science.abc7807

PHOTO: CHUCK PAINTER/STANFORD NEWS SERVICE

Published by AAAS

Circles on the grounds of San Francisco's Dolores Park are designed to limit the spread of SARS-CoV-2 by encouraging social distancing.

PHOTO: NOAH BERGER/AP PHOTO

SOCIAL SCIENCE: COVID-19

Which interventions work best in a pandemic?

We can exploit randomized controlled trials, compartmental models, and spillovers

By **Johannes Haushofer**[1,2,3,4,5] *and* **C. Jessica E. Metcalf**[6]

The only approaches currently available to reduce transmission of the novel coronavirus severe acute respiratory syndrome–coronavirus 2 (SARS-CoV-2) are behavioral: handwashing, cough and sneeze etiquette, and, above all, social distancing. Policymakers have a variety of tools to enable these "nonpharmaceutical interventions" (NPIs), ranging from simple encouragement and recommendations to full-on regulation and sanctions. However, these interventions are often used without rigorous empirical evidence: They make sense in theory, and mathematical models can be used to predict their likely impact (*1*, *2*), but with different policies being tried in different places—often in complicated combinations and without systematic, built-in evaluation—we cannot confidently attribute any given reduction in transmission to a specific policy.

Because many of these interventions differ from each other in terms of their economic and psychological cost—ranging from very inexpensive, in the case of interventions based on behavioral economics and psychology, to extremely costly, in the case of school and business closures—it is crucial to identify the interventions that most reduce transmission at the lowest economic and psychological cost. Randomized controlled trials (RCTs) are one of several methods that can be used for this purpose but surprisingly have received little attention in the current pandemic, despite a long history in epidemiology and social science. We describe how RCTs for NPIs can be practically and ethically implemented in a pandemic, how compartmental models from infectious disease epidemiology can be used to minimize measurement requirements, and how to control for spillover effects and harness their benefits.

JUSTIFIABLE RCTS

How can RCTs be practically and ethically conducted in a pandemic? In a typical RCT, a subset of randomly chosen individuals or regions receives an intervention, and a randomly chosen control group receives no intervention or a different intervention. The random assignment ensures that any later differences between the groups can be attributed to the intervention. During an outbreak, policy-makers must decide which interventions to impose when, and when to loosen them again. It will rarely be feasible in this context to omit individuals or regions entirely. However, policy-makers can use systematic timing of such interventions to both protect the population and understand the impact of the intervention. For example, when experts begin to think that measures can be loosened, this can be done gradually, so that evaluation is possible: A subset of randomly chosen locations (such as counties or municipalities) begins, and others gradually follow suit. Comparison of the "early" to the "late" regions makes it possible to estimate the effects of the intervention.

This "phase-in" or "stepped-wedge" approach can be used at any point during the

[1]Department of Psychology and Woodrow Wilson School of Public Affairs, Princeton University, Princeton, NJ, USA. [2]National Bureau of Economic Research, Cambridge, MA, USA. [3]Busara Center for Behavioral Economics, Nairobi, Kenya. [4]Center for Global Development, Washington, DC, USA. [5]Max Planck Institute for Collective Goods, Bonn, Germany. [6]Department of Ecology and Evolutionary Biology and Woodrow Wilson School of Public Affairs, Princeton University, Princeton, NJ, USA. Email: cmetcalf@princeton.edu; haushofer@princeton.edu

Published by AAAS

pandemic. At the beginning, protective measures can begin early in some areas and somewhat later in others. During the pandemic, periods of loosened measures may be necessary to restore a sense of normality and keep essential services working, or measures may have to be tightened to limit further spread of the virus; these periods can also be systematically timed to evaluate their impact. In extended versions, different interventions can be tested against each other, and different locations can tighten or loosen different subsets of restrictions; for example, schools could be opened back up, whereas nonessential businesses remain closed.

Governments and organizations could work with scientists to choose an experimental design, implement and keep track of the treatment assignment, and measure outcomes. Studies of this kind can now often be done in nimble and practicable ways, reducing the oversight and time burden on implementing partners. Interventions could range from messaging campaigns to promote social distancing to laws and regulations. Where full randomization (without phase-in) is possible, this may be desirable to increase statistical power (*3*).

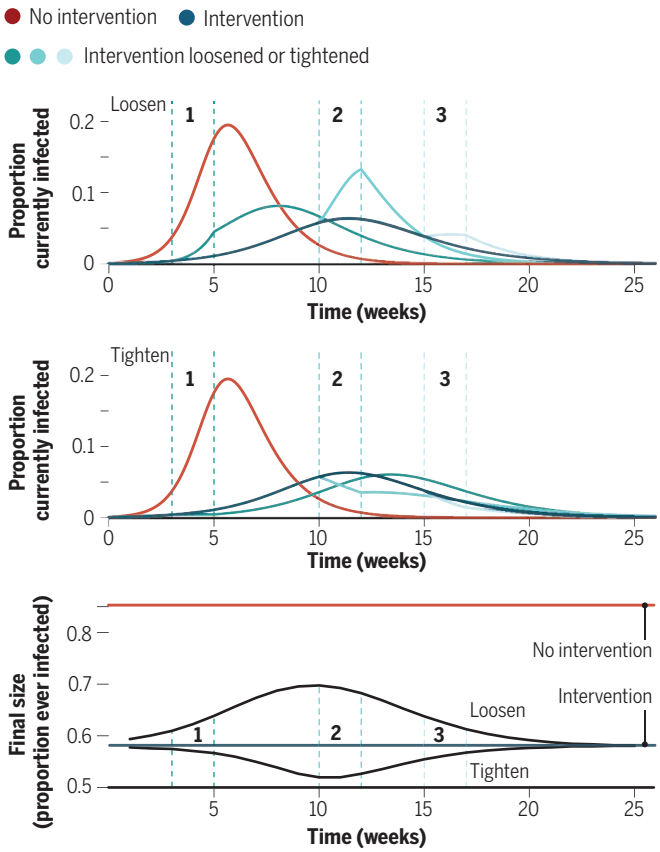
RCTs are, of course, not the only method for estimating the impact of NPIs. Where randomization is not feasible, the "natural experiments" created by some policies can be exploited, such as quasi-arbitrary cutoffs (for example, the reopening of stores below a certain square footage). Observational studies, often integrated with mathematical models have also contributed important insights.

Great care must be exercised to make RCTs ethical. Several considerations are relevant: The approach may be ethically justifiable because there are two sources of uncertainty around most interventions. For any intervention, it may be uncertain whether its benefits in terms of reducing disease transmission exceed its economic and psychological costs or how these costs and benefits relate to those of other interventions. At the same time, it is difficult to identify a single "correct" moment to loosen or tighten protective measures, as illustrated by ongoing policy debates. Thus, equipoise may be satisfied in terms of costs, benefits, and timing. Policy-makers are therefore neither knowingly withholding a beneficial intervention from constituents nor knowingly imposing a harmful one. This uncertainty is likely to make staggered tightening or loosening of an intervention more acceptable to the public.

Further, the phase-in or stepped-wedge approach may be ethically justifiable because individuals in both control and treatment groups eventually experience the costs and benefits of any intervention. In addition, even short periods of tightening or loosening can be used to determine the impact of mitigation measures, minimizing the burden on whichever group experiences the smaller benefits. A powerful illustration of the ethical acceptability of this phase-in approach among both scientists and the public is its use in RCTs of vaccines, even for highly lethal pathogens such as Ebola (*4*).

Testing interventions during an outbreak

Time course of infection in the absence of an intervention (red) and with an intervention (blue) that is either loosened (top) or tightened (middle) for 2 weeks. The number of cases at the end of different tightening or loosening windows (bounds indicated with dashed vertical lines) in regions where the intervention was tightened or loosened compared with regions where it remained in place provides a measure of the relative change in transmission associated with the intervention. Final size (bottom) is affected by the time point at which the 2-week tightening or loosening of the intervention is initiated (*5*). See supplementary materials for details.



MODELS TO GUIDE DATA COLLECTION

Careful measurement of outcomes is crucial for this approach to succeed. In particular, it is essential to understand the impact of any given intervention on the full epidemic trajectory [see supplementary materials (SM)]. However, the measurement requirements can be simplified if data collection and analysis are guided by compartmental models from infectious disease epidemiology. The time course of infections is affected in a SIR model (reflecting the three possible states of an individual in the community: susceptible, infectious, or recovered) when one group of locations (such as counties or districts) loosens or tightens the intervention for 2 weeks while another maintains the status quo (see the figure) (*5*). Crucially, because the SIR model describes the entire trajectory of an outbreak using only two parameters, very few measurements are required to estimate them. In particular, using only estimates of the number of infections at the end of the intervention in treatment and control regions, we can estimate how much a given intervention reduces transmission relative to no intervention (see SM). In addition, this difference allows policy-makers to determine which of several interventions reduced transmission the most and by how much. If additional information about the number of infections at the beginning of the intervention is available, we can further estimate whether transmission has been sufficiently reduced that the outbreak is shrinking (corresponding to an effective reproductive number below 1).

Insights from epidemiology can also be used to address several additional questions: In addition to learning how much an intervention changes the transmission rate, policy-makers may also want to know how different interventions affect the "final size" of the pandemic—what share of the population will have been infected in total when the pandemic has died down. Also, they may want to understand how a single intervention might perform if it were deployed at different time points during the pandemic (for example, early versus late) but can only test it once. Additionally, they may

GRAPHIC: X. LIU/*SCIENCE*

Published by AAAS

want to directly compare the effectiveness of two interventions, despite them having been deployed at different times, because it may not always be possible to time periods of tightening and loosening precisely.

In the stylized model, all of these estimates can be derived from adding a single measurement at one time point to those described above—namely, the number of susceptibles [measured, for example, with serology (*6*)] (see SM). Of course, the available capacity for polymerase chain reaction and serology needs to be able to support such studies, but testing capacity is growing around the world, suggesting feasibility.

An important caveat to reducing the measurement requirements is that the above approach leans on the assumptions of a relatively simple SIR model; in particular, both the transmission rate and the impact of an intervention on this rate are assumed to remain constant throughout the pandemic. However, it is straightforward to extend the model to accommodate inherent variation in transmission through time, and complex treatment effects that may be a signature of NPIs, including decay over time (for example, fatigue from a lockdown or fading response to a messaging campaign), persistence (for example, hygiene behaviors such as handwashing which turn into a habit), or intensification over time (such as messaging campaigns that "go viral"). In such cases, the measurement requirements will increase to identify the additional parameters (such as decay) contained in the extended model. Similarly, the basic compartmental model can be extended [to SIRS (susceptible-infected-recovered-susceptible), SEIR (susceptible-exposed-infectious-recovered), or age-structured models, for example] to reflect additional features of the transmission process (duration of immunity, latent period, or variable contact patterns over age) or the intervention (for example, if it targets specific age groups).

Thus, the effects of interventions on disease transmission can be estimated with the help of epidemiological models. However, the economic and psychological costs and benefits of such interventions are equally important. Reducing the number of measurements by leveraging the SIR model is not possible for these outcomes, about which the model makes no predictions and whose time course need not follow that of infections. For example, a "successful" intervention that reduces the risk of overburdening the health system will have the effect of spreading the infections over time. This implies that the desirable behaviors induced by any intervention have to be maintained for longer to outlast the duration of the pandemic. This may impose psychological and economic costs on the population that are larger than those that would be incurred in a more temporally condensed pandemic. In the absence of a model, these effects can only be captured with careful measurement over time.

SPILLOVER EFFECTS

Interventions delivered to some regions or individuals but not others are likely to nevertheless affect those who were not targeted. Such so-called "spillover" effects present both a challenge and an opportunity in evaluating the impact of NPIs. The opportunity is that such spillovers can generate strongly increasing returns to intervention coverage in terms of individual protection (*7*); they can therefore be harnessed to maximize the effects of a given intervention. For example, consider a hypothetical intervention that reduces the size of a pandemic by 15% when it is delivered to 20% of a community. Because of the nonlinear dynamics of infection that arise from depletion of the number of those susceptible to infection, increasing the coverage to 60% may generate a greater-than-proportional reduction in pandemic size of 56%.

At the same time, such spillovers pose a challenge to the estimation of treatment effects. However, standard trial designs are available to enable measurement of spillovers (*8*–*10*). In particular, nonlinear returns to saturation (the share of the population exposed to an intervention) can be integrated into tests of interventions by creating variation in spatial saturation of intervention delivery. For example, groups of 15 locations might be randomized to a "low saturation" condition in which a third of locations are treated with an intervention—for example, the distribution of face masks or hand sanitizer, or opening or closing of parks or schools—or to a "high saturation" condition, in which two-thirds of locations are treated. Such studies have to be relatively large scale to achieve adequate statistical power; power calculations are therefore important, and using more than two or three levels of saturation may not be practicable.

Because spatial spillovers may occur at different spatial scales, causal inference methods that flexibly allow for such complications have to be used. Data on the source of spillovers, such as the commuting patterns of essential workers, can help identify relevant spatial scales. The feasibility of this approach in terms of both statistical power and causal inference in the presence of spillovers of unknown spatial dimensions has been suggested by recent large-scale studies on the general equilibrium effects of economic interventions (*11*). Thus, tests of interventions to combat COVID-19 should take advantage of, and measure, these nonlinear effects of saturation.

NPIs can be rigorously tested by using randomization without compromising scientific and ethical standards. Although this approach will require more time than generating projections from observational methods and mathematical models, the benefits in terms of accuracy could be considerable. If policy-makers and scientists combine insights from infectious disease epidemiology with carefully and ethically designed impact evaluation, alongside other empirical and theoretical methods for studying impact (*12*–*14*), they will have a powerful tool for reducing the human health, societal, and economic costs in the SARS-CoV-2 pandemic and in pandemics in general. ■

REFERENCES AND NOTES

1. N. Ferguson *et al*., "Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand," Working paper, Imperial College COVID-19 Response Team, London, UK, 16 March 2020.
2. T. Rawson *et al*., medRxiv 20084517 [Preprint] 2 May 2020; https://doi.org/10.1101/2020.04.29.20084517
3. L. Kennedy-Shaffer, M. Lipsitch, medRXiv 20087429 [Preprint] 6 May 2020; https://doi.org/10.1101/2020.05.01.20087429
4. A. M. Henao-Restrepo *et al*., Lancet **389**, 505 (2017). doi:10.1016/S0140-6736(16)32621-6 Medline
5. F. Di Lauroy, I. Z. Kissy, J. C. Miller, "The timing of one-shot interventions for epidemic control," Working paper, La Trobe University, Melbourne, Australia.
6. C. J. Metcalf *et al*., *Lancet* **388**, 728 (2016). doi:10.1016/S0140-6736(16)30164-7 Medline
7. R. M. Anderson, R. M. May, Infectious Diseases of Humans: Dynamics and Control (Oxford Univ. Press, 1992).
8. J. Benjamin-Chung *et al*., *Int. J. Epidemiol*. **47**, 332 (2018); doi:10.1093/ije/dyx201 Medline
9. M. E. Halloran, C. J. Struchiner, *Epidemiology* **2**, 331 (1991); doi:10.1097/00001648-199109000-00004 Medline
10. E. Miguel, M. Kremer, Econometrica **72**, 159 (2004). doi:10.1111/j.1468-0262.2004.00481.x
11. D. Egger *et al*., "General equilibrium effects of unconditional cash transfers: Experimental evidence from Kenya," Working paper26600, NBER, Cambridge, MA, 21 November 2019.
12. H. Fang, L. Wang, Y. Yang, "Human mobility restrictions and the spread of the novel coronavirus (2019-nCoV) in China," Working paper 26906, National Bureau of Economic Research (NBER), Cambridge, MA, March 2020.
13. S. Flaxman *et al*., "Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries, Working paper, Imperial College COVID-19 Response Team, London, UK, 30 March 2020.
14. H. Tian *et al*., *Science* **368**, 638 (2020); doi:10.1126/science.abb6105 Medline

ACKNOWLEDGMENTS

We thank D. Björkegren, A. Chandrasekhar, C. de Chaisemartin, J. de Quidt, B. Grenfell, R. Hussam, S. Jayachandran, D. Strömberg, and anonymous referees for helpful comments and suggestions.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6495/1063/suppl/DC1

Published online 21 May 2020
10.1126/science.abb6144

Published by AAAS

SCIENCE AND REGULATION

Filling gaps in science exposes gaps in chemical regulation

Examination of U.S. and EU regulatory systems raises more questions than answers

By **Steve C. Gold**[1] *and* **Wendy E. Wagner**[2]

The regulation of chemicals should protect public health and the environment from undue risk of harm, should promote the development and use of safer alternatives to more hazardous chemicals, and should provide the public with sufficient information to understand how well chemical risks are being managed. How well are these goals being achieved? The regulatory system in the United States has been derided as dysfunctional (*1*), even with major amendments enacted in 2016 (*2*) that some supposed would bring the U.S. program closer to the European Union's REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals) Regulation. To build on the literature that documents the shortcomings of chemical regulation (*1*, *2*), we take as a convenient example the compounds described in new research by Washington *et al.* on page 1103 (*3*). Washington *et al.* report the unexpected presence in environmental samples of chloroperfluoropolyether carboxylate compounds (ClPFPECAs), apparently used as a substitute for other perfluoroalkyl substances (PFASs) that had raised environmental concerns (*3*). Attempting to trace these compounds through the regulatory regime raises more questions than answers, revealing the structural limits of existing regulation. These limits apply not only to this particular case but to myriad chemicals. How much confidence do regulatory systems give the public that substitute chemicals are safer than the substances they replace? Not nearly as much as one would like.

BIRTH OF A NEW CHEMICAL

The compounds found by Washington *et al.* apparently were first reported in the literature as part of the search for "environmentally friendly" replacements for chlorofluorocarbons and hydrofluorocarbons (*4*). They appear to have found a use, however, in the manufacture of fluorinated polymers with nonstick properties for use, for example, in cookware (*5*, *6*). As such, they helped answer an urgent desire to eliminate the use and release of perfluorooctanoic acid (PFOA) and its precursors and homologs (*3*), which had been widely dispersed in the environment as a result of the production of other substances with similar properties [see supplementary materials (SM)]. Amid growing evidence that PFOA and similar compounds were highly persistent, bioaccumulative, and potentially toxic, as well as a spate of private lawsuits and increasing state and federal government scrutiny (see SM), eight companies voluntarily agreed "to work toward eliminating PFOA from emissions and in product content" in the United States by 2015 (*7*). By the time of this agreement between the companies and the U.S. Environmental Protection Agency (EPA), the phase-out of PFOA was already well under way informally (see SM). EPA's PFOA Stewardship Program required the companies to submit annual reports of their progress, although, as a voluntary agreement, it had no enforcement mechanism (*7*). And because the program focused only on reducing and eventually eliminating PFOA emissions and PFOA in product content, it did not in any way address any substitute compounds the companies might develop, use, generate, or release in lieu of PFOA. Toxicity testing and reporting for substitute compounds were simply outside the scope of the PFOA Stewardship Program.

Toxicity testing has revealed reasons for concern about some of these alternatives (*8*, *9*), leading regulatory authorities in at least one U.S. state, New Jersey, to search for replacement as well as legacy PFASs in the environment (*3*, *9*). The report by Washington *et al.*—whose authors are researchers affiliated with EPA and the New Jersey Department of Environmental Protection (DEP)—grew out of that effort. Given the nature of the PFOA Stewardship Program, as well as other features of U.S. chemical regulation that we describe below, it makes sense that the ClPFPECAs the investigators found were previously unknown to them.

Washington *et al.* detected ClPFPECAs in every soil sample they tested from New Jersey, as well as in a stored soil sample taken during earlier research at a location more than 400 km away (*3*). They concluded that their data "strongly suggest atmospheric release" of these compounds from a New Jersey facility of a Solvay S.A. business unit (*3*). They also found these compounds in a previously taken water sample from a river in Italy, which they used to help confirm their identification of the compounds (*3*). The water sample result agreed with prior findings by other researchers who found ClPFPECAs in the same river (*6*). In light of the fairly widespread detection of these compounds in environmental samples, it is reasonable that the researchers suggest that further investigation of these compounds' environmental fate, transport, and degradation is warranted, and that "investigation of whether these ClPFPECAs might be toxic is prudent" (*3*).

PUBLIC TOXICITY INFORMATION

But what can we say right now, based on publicly available information, about "whether these ClPFPECAs might be toxic"? One might think that government regulators could provide some answers for the public. As it turns out, however, trying to trace these compounds through the U.S. and European regulatory systems yields frustratingly unsatisfying answers and reveals a dearth of publicly available research to support such answers as may exist.

In the United States, regulators in California have expressed serious concerns that perfluoroether carboxylic acids—a class that includes the ClPFPECAs (see SM)—"may have similar or higher toxic potency than the longer-chain PFAAs [perfluoralkyl acids] they are replacing" [(*10*), p. 38] and are similarly "recalcitrant to degradation and extremely persistent in the environment" [(*10*), p. 10]. The agency also stated that perfluoropolyethers, in general, may contain as impurities, or upon combustion may release, PFAAs that are persistent, bioaccumulative, and potentially toxic [(*10*), p. 12]. Yet the document expressing these concerns focused primarily on GenX, another chemical in the class, and did not discuss the properties or potential toxicity of ClPFPECAs in particular (*10*). GenX and other PFAS chemicals have also been detected in environmental media at levels that raise regulatory concern (*9*, *11*).

As for Europe, a search by Chemical Abstracts Service (CAS) number 329238-24-6 reveals that the European Chemicals Agency (ECHA) requires classification and labeling of these compounds under the European Classification, Labelling and Packaging (CLP) Regulation. On ECHA's chemical home page, the ClPFPECAs trigger

[1]Rutgers Law School, Rutgers University, Newark, NJ, USA. [2]University of Texas School of Law, Austin, TX, USA. Email: stgold@law.rutgers.edu

Published by AAAS

five of ECHA's hazard codes: The substance is fatal if swallowed, is fatal in contact with skin (i.e., acutely toxic), causes severe skin burns and eye damage, causes liver damage through prolonged or repeated exposure, and is toxic to aquatic life with long-lasting effects. For 12 other hazard categories, the entries simply read "data lacking." No research to support any of these classifications is cited on or publicly available through the ECHA website.

On the other hand, in 2010 the European Food Safety Authority (EFSA) approved ClPFPECAs for use in the polymerization of anti-stick coatings for repeated-use foodware, subject to restrictions on process quantities and temperatures (*5*). EFSA found the chemical safe for these applications. EFSA concluded that the substance passed a bacterial gene mutation test, an in vitro mammalian cell gene mutation test, and an in vitro mammalian cell chromosome aberration test. The sole source cited was a "dossier" provided by the manufacturer, without any studies that the dossier may have included or cited (*5*). In response to our request for further information under the EU's regulation concerning public access to documents, EFSA provided titles and dates for the three unpublished studies listed in the dossier, but the names and affiliations of the authors were withheld as personal data. As we went to press on 29 May, we were still awaiting a response from EFSA to our further request for access to the full dossier, because the studies cannot be found in the public domain (see SM).

Thus, in our search for toxicity information, we found more questions than answers. There are dire but generalized concerns of the state of California; five hazard classifications listed by ECHA, with "data lacking" for 12 others; and a benign assessment by EFSA, based on limited and unpublished data, with respect to polymerization of foodware coatings. We also checked an enormous chemical-toxicological database maintained by EPA (Comptox), which includes information on some 875,000 chemical substances (far more than the universe of regulated substances). We found an entry for the ClPFPECAs that includes an "executive summary" of the available data in 32 categories of information relevant to toxicity or hazard; no data or values are shown for any of the 32 categories (see SM). Looking beyond government agency websites, we also found no toxicological studies of ClPFPECAs in a literature search.

THE REGULATORY STORY

This scientific inconclusiveness leads to the next obvious set of questions: How could ClPFPECAs appear to be so unstudied and unaccounted for? After all, ClPFPECAs share at least some properties with PFASs that are being phased out (*6*, *9*, *10*). Society would presumably want to know that substitute chemicals like these are not worse than the chemicals they replace. Shifting focus to the regulatory system, however, yields few definitive answers about the oversight of ClPFPECAs, although it does offer several lessons about the design of our regulatory programs more generally.

The first lesson, just touched on above, is that while regulatory attention is focused on eliminating high-profile chemical risks, less effort appears to be dedicated to analyzing the safety of substitute chemicals used to replace them (*12*, *13*). A number of scientists have raised general concerns about the need for rigorous comparative assessments of replacement chemicals, particularly within the PFAS family (*12*, *13*). This type of comparative analysis seems particularly appropriate in light of the voluntary phase-out of PFOA, all the more so because EPA has identified about 500 PFAS chemicals sold in U.S. commerce, out of a larger, global list of thousands of such compounds (see SM). However, assessments for PFAS chemicals appear to have been conducted—at best—on an ad hoc basis and primarily through negotiated agreements. The resulting, publicly available research on PFAS chemicals is quite limited. The state of New Jersey reports that out of a list of 900 PFAS chemicals, only 200 chemicals have toxicity data available at all, and even that research is incomplete (*9*). EPA has now instituted more uniform comparative assessment procedures for some PFAS chemicals, but these recently revised procedures apply only to new polyfluoroalkyl chemicals or uses produced after 2015 (see SM). It is of course possible that despite the regulatory vacuum, a comparative analysis was nonetheless performed internally by the manufacturer. If that analysis exists, however, it does not appear to be publicly available.

What about oversight of ClPFPECAs under the U.S. Toxic Substances Control Act (TSCA)? TSCA is a regulatory program originally enacted in 1976 to prevent unreasonable risks caused by chemicals. Under TSCA, makers of "new" chemicals (developed after 1976) must submit a premanufacture notification to EPA (*1*). And the ClPFPECAs certainly seem, on their face, to represent a distinctly "new chemical" developed after 1976 (*4*). A search through EPA's TSCA inventory, however, provides more unsolved mysteries: The ClPFPECA family is not listed (by CAS number) in EPA's public inventory of more than 40,000 registered chemicals (see SM).

It is theoretically possible that the manufacturer simply violated EPA's registration requirements under TSCA, but there are several more likely explanations for why the ClPFPECAs are not listed in EPA's inventory. One is that although the law generally requires premanufacture notification of "new" chemicals, there are multiple exemptions from this registration requirement. It is unclear whether the ClPFPECAs satisfied any of these exemptions, which allow manufacturers to avoid submitting a premanufacture notification, for example, on new chemicals that are long-chain polymers or that are only impurities. But we cannot know for sure, because EPA generally does not provide public tracking of the manufacturers' use of these various exemptions (see SM).

Alternatively, it is possible that the ClPFPECAs are in fact tracked under TSCA, but the chemical is not listed in the public database because the chemical structure was classified by the manufacturer as a protected trade secret (called CBI, for "confidential business information") and removed from public view. Currently, more than 140 unidentified PFAS chemicals in U.S. commerce are classified as CBI, according to EPA (see SM). Once information is stamped by a manufacturer as CBI, only cleared government staff can view the files (*14*). Even the generic names of the CBI chemicals are not made public (*14*). Yet a CBI classification seems inapplicable to the ClPFPECAs because its chemical structure has been published (*4*). In practice, though, a CBI claim remains legally in place until either the manufacturer or EPA officially "declassifies" the claim (*14*). Because more than 10,000 chemicals in the TSCA inventory are classified as CBI (*14*), Congress in 2016 required EPA to review and, if warranted, declassify a subset of CBI chemi-

Where are the data?

The EPA Comptox chemical-toxicological database includes an entry for the ClPFPECAs, but indicates that no data or values are available for any of the categories of information. See (*16*).

- ✗ Quantitative Risk Assessment Values
- ✗ Quantitative Hazard Values
- ✗ Cancer Information
- ✗ Reproductive Toxicology
- ✗ Chronic Toxicology
- ✗ Subchronic Toxicology
- ✗ Developmental Toxicology
- ✗ Acute Toxicology
- ✗ Subacute Toxicology
- ✗ Neurotoxicology
- ✗ Endocrine System
- ✗ Absorption, Distribution, Metabolism, and Elimination
- ✗ Fate and Transport
- ✗ Exposure
- ✗ Adverse Outcome Pathway Information
- ✗ Water Quality
- ✗ Air Quality
- ✗ Occupational Exposure

Published by AAAS

cals. The agency is still working on this assignment, with about 2000 chemicals left to review (see SM). ClPFPECAs may well be in this queue, but confirming whether they are is not easy; it likely requires a formal request under the federal Freedom of Information Act. We can therefore extract a second lesson about the regulatory system: Some chemicals may fall through the cracks in the public tracking system in the United States, not because they are adequately assessed for toxicity but for other reasons.

Yet just because ClPFPECAs may not be in public view does not mean that the chemicals are out of range of U.S. regulators. Under TSCA, it is possible that even if the product is protected as a trade secret, EPA is actively overseeing the ClPFPECAs. Under this "extensive regulation" scenario, EPA could ask for more testing or other negotiated restrictions as a condition to approving ClPFPECAs under TSCA. Or the agency may be satisfied—after conducting its own analysis—that the risks of ClPFPECAs are not unreasonable. EPA might even impose new restrictions on the manufacture of ClPFPECAs over time if it learns of new "adverse effects," which manufacturers are required by law to report.

Conversely, even if ClPFPECAs were subject to premanufacture regulatory review, it is also possible that EPA decided to pass ClPFPECAs into commerce without much, or any, testing or analysis. Only about 15% of premanufacture notifications for new chemicals submitted to EPA include any health or safety test data at all [(*15*), p. 11], and EPA's own statistics show that only 10% of the new chemicals entering commerce between 1979 and 2016 involved restrictions or testing orders (see SM). In the case of the ClPFPECAs, then, EPA may have been concerned that there was insufficient scientific evidence available at the time to support an order demanding more testing. Under the pre-2016 law, EPA was required, within a brief 90-day period, to produce some evidence of potential risk as a predicate to ordering additional toxicity tests from the manufacturers. This catch-22 resulted in a paucity of testing orders (*1*, *2*). [Congress removed this legal impediment in the 2016 TSCA amendments, but that came too late for the ClPFPECAs (*2*).]

We thus don't know much, if anything, about the regulatory oversight of ClPFPECAs in the United States. This mystery provides yet a third lesson about chemical regulation: For the 40,000-plus chemicals in commerce, the burden of chemical assessment rests almost entirely on a small group of EPA regulators (*1*). As a result, some, perhaps many, chemicals likely fall through the cracks. Indeed, despite the amendments to TSCA in 2016, chemical manufacturers are still not required to anticipatorily test or assess their chemicals as a condition to marketing in the United States (*1*, *2*). Instead, it is the regulators who bear responsibility for identifying the most hazardous chemicals, identifying the relevant scientific literature bearing on toxicity, ordering new tests, and synthesizing and analyzing the available information bearing on the safety of individual chemicals (not to mention determining proper policy responses) (*1*, *2*). Manufacturers' primary role is to serve as respondents. In this role, they not only provide the information that EPA requests, but also have the right to lodge extensive and critical comments on the agency's work. Manufacturers can also sue the agency in court, arguing that some aspect of the agency's analysis might be arbitrary (see SM). And the new powers EPA gained from the 2016 amendments—such as the ability to require additional testing and prioritize chemicals of concern—came with substantial procedural impediments (*2*).

What about Europe? ClPFPECAs are officially registered in two EU regulatory programs. As mentioned above, EFSA approved ClPFPECAs in the manufacture of nonstick coating products (*5*). And ClPFPECAs are listed—along with five hazard classifications—in the EU's notification (CLP) database (see SM). In neither case, however, is the supporting research behind these regulatory findings readily available to the public.

Additionally, in 2005, the European Union implemented REACH, a more aggressive chemical regulatory program than the U.S. TSCA program. Unlike TSCA, REACH requires manufacturers to conduct a comprehensive literature search on the toxicity of each of their chemicals sold above threshold quantities in the European Union. If the existing scientific information proves incomplete under REACH standards, manufacturers are also required to conduct additional toxicity testing (*1*). Yet a search by CAS number returns no results for ClPFPECAs on the European Chemical Agency's website that lists registered REACH chemicals. Perhaps ClPFPECAs are produced in low enough quantities (less than 1 tonne/year) to be exempted from REACH, or perhaps ClPFPECAs satisfy other REACH exemptions, such as those governing impurities or polymers as defined under REACH. Publicly available information again does not resolve this question (see SM).

A LONG WAY TO GO

Public health and environmental concerns led to the phase-out of PFOA and its closest chemical relatives. It seems obvious that in such a scenario, society should want to be reasonably certain that the replacement chemical "cure" is not worse than the phased-out chemical "disease." Yet the researchers from EPA and New Jersey DEP found substitute polyfluorinated compounds, ClPFPECAs, in their environmental samples (*3*). At present, there is little in the public scientific record to indicate whether the environmental dissemination of these substitute PFASs is benign or harmful, and if harmful, how harmful. Our examination of the U.S. and European regulatory programs raises more questions than answers about the extent to which ClPFPECAs are being tracked, studied, and regulated. Certainly the European Union has made more progress than has the United States in this regard (*5*). Still, the toxicological mysteries of ClPFPECAs—and thousands of other potentially toxic chemicals that are regulated (or perhaps not regulated) in ways that remain effectively inscrutable—suggest that we have a long way to go in designing effective and accountable chemical regulation, particularly in the United States. ■

REFERENCES AND NOTES

1. J. G. Applegate, *Ecol. Law Q.* **35**, 721 (2008).
2. V. J. Watnick, *Harv. Envtl. L. Rev* **43**, 373 (2019).
3. J. Washington *et al.*, *Science* **368**, 1103 (2020).
4. C. Tonelli, A. Di Meo, R. Picozzi, *J. Fluor. Chem.* **128**, 46 (2007).
5. EFSA Panel on Food Contact Materials Enzymes, Flavourings and Processing Aids, *EFSA J.* **8**, 1519 (2010).
6. M. Mazzoni *et al.*, *Norman Bull*. No. 4, p. 5 (2015).
7. EPA docket on PFOA voluntary stewardship program, docket number EPA-HQ-OPPT-2006-0621; https://www.regulations.gov/document?D=EPA-HQ-OPPT-2006-0621-0005
8. Y. Wang *et al.*, *Ecotoxicol. Environ. Saf.* **182**, 109402 (2019).
9. New Jersey Department of Environmental Protection, "Comments on USEPA Draft Human Health Toxicity Assessments for GenX and PFBS" (2019); https://downloads.regulations.gov/EPA-HQ-OW-2018-0614-0020/attachment_1.pdf
10. California Department of Toxic Substances Control, *Product-Chemical Profile for Carpets and Rugs Containing Perfluoroalkyl or Polyfluoroalkyl Substances, Final Version* (2019); https://dtsc.ca.gov/wp-content/uploads/sites/31/2020/02/Final_Product-Chemical_Profile_Carpets_Rugs_PFASs_a.pdf
11. M. Goot, *North Carol. Law Rev*. **98**, 629 (2020).
12. P. Grandjean, *Environ. Health* **17**, 62 (2018).
13. M. M. Jacobs, T. F. Malloy, J. A. Tickner, S. Edwards, *Environ. Health Perspect*. **124**, 265 (2016).
14. Environmental Working Group, *Off the Books II: More Secret Chemicals* (2016); www.ewg.org/research/off-the-books-ii-more-secret-chemicals/introduction#_edn1
15. U.S. Government Accountability Office, *Chemical Regulation: Options Exist to Improve EPA's Ability to Assess Health Risks and Manage Its Chemical Review Program* (2005); www.gao.gov/new.items/d05458.pdf
16. https://comptox.epa.gov/dashboard/dsstoxdb/results?search=329238-24-6#exec_sum

ACKNOWLEDGMENTS

We thank anonymous reviewers and A. Gore, J. Scott, and V. Watnick for valuable feedback and comments. S.C.G. and W.E.W. declare no competing interests.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6495/1066/suppl/DC1

10.1126/science.abc1250

Published by AAAS

BOOKS *et al.*

A ship traverses the Bering Sea near Russia's Chukchi Peninsula.

PHOTO: 1441147448/SHUTTERSTOCK

ECOLOGY

Colonialism and its consequences

A ravaged Arctic ecosystem serves as a warning of the perils of human advancement

By **Mary Ellen Hannibal**

The area known as Beringia sits atop the world, straddling Asia and North America. A mere 50 miles of water separate the Chukchi Peninsula in Russia and the Seward Peninsula in the United States. Terrestrial plant life is relatively impoverished here, but the Bering Strait compensates for the land's parsimony and more.

In *Floating Coast*, Bathsheba Demuth tells the story of this singular place beginning in the mid-1800s. Empire-making was under way then, both by the United States and by Russia, and the lifeways of native Iñupiat, Yupik, and Chuckchi peoples were about to be severely curtailed. So too were the natural histories of the species that contributed to making the region.

America was earnestly building its sovereign might, and thus its economy, during this period. Having exhausted species off both the Atlantic and Pacific coasts, whalers looking for new quarry came to Beringia to hunt bowheads. Whales were integral to just about every machine and product made at the time. Whale blubber was used for lubricating sewing machines and cotton gins. Whalebone (baleen) lent structure to umbrellas, fishing rods, and mattresses. Whale tallow was refined into soap and became a base for perfume. Most of all, whale oil was used to fuel indoor lighting.

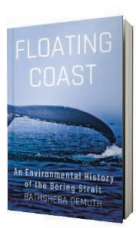
The ever-ratcheting appetite for whales drove them to near extinction in a few short decades. The indigenous peoples whose lives depended on whales lost both their source of sustenance and the culture around which their identities had been constructed for millennia.

With whale populations depleted, industrial appetites turned to walruses, although it took 250 of the blubbery beasts to measure up to a single bowhead. "Far from Beringia," Demuth writes, "walruses were a minuscule part of imagining a prosperous, mechanized future: as fan belts on power looms or grease in factory cogs…." Wholesale depletion of walruses continued well into the early 1900s, as their blubber was transformed into nitroglycerin in service of World War I.

While Americans denuded Arctic waters, Russians bristled at what they considered theft of their natural resources. The denizens of the waters between the two sovereign nations did not adhere to any putative human boundary.

"The problem of whaling, for the United States and Imperial Russia, was not so much that it killed too many whales, but that it brought commerce while failing at civilization," writes Demuth. The United States bemoaned profits that were being spent on guns and alcohol, contributing not to lawfulness but its opposite. The Russians saw those profits as going to the wrong people (Americans). The solution that was negotiated was to establish boundaries, to "enclose" the cyclical peregrinations of whales, and subsequently those of walruses, reindeer, and fox, in an attempt to systematize harvests and make their yields more reliable. Bringing little ecological knowledge to bear on this effort led to both market and population crashes: "…while dollar value might make fur seem the same as blubber or ivory, the rhythms that governed their creation were not equivalent; the market's habit of supplanting one desire for another did not apprehend how a fox would never live on the same time as a walrus."

As the Russian Revolution transformed into bolshevism, and from thence to communism, the question of how exactly to govern territory according to a Marxist ideal brought new pressures to Beringian land, sea, and native peoples. The cycles of nature notwithstanding, Stalinism in particular sought to mechanize production according to its own time frame, not only at the expense of the animals involved but by way of cruelty to the Russian people it employed.



Floating Coast: An Environmental History of the Bering Strait
Bathsheba Demuth
Norton, 2019. 416 pp.

Once nonliving resources in the form of gold and eventually tin and oil became the focus of colonial harvest, the question of how to enclose Beringia became expressly about sovereignty. "In Alaska and Russia at the turn of the twentieth century, the politics of mining had to do with who should rightfully benefit from gold: the individual proprietor, the corporation, the tsar, or the collective."

In Demuth's masterful narrative, the very long time frame of energy production crashes disastrously into the ever more expedient ways humanity purposes that energy. Global warming, the result of quickly burning organic material that took eons to accumulate, is the apotheosis of colonial strategy. *Floating Coast* is eloquent testimony to how this strategy is not working. ■

The reviewer is the author of *Citizen Scientist: Searching for Heroes and Hope in an Age of Extinction* (The Experiment, 2016). Email: maryellenhannibal@gmail.com

10.1126/science.abc2285

Published by AAAS

SCIENCE LIVES

Where logic meets emotion

A computer scientist gets candid about her quest to bring empathy to artificial intelligence

By **Vijaysree Venkatraman**

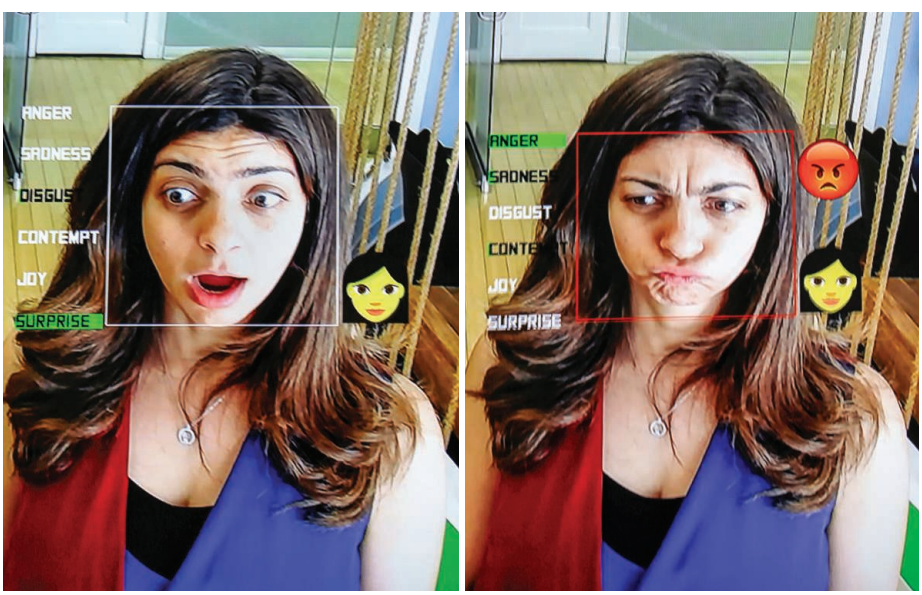
As she sat in a taxi headed to Cairo International Airport in September 2001, Rana el Kaliouby remembers thinking, "*Am I really going through with this?*" A married woman and hijab-wearing Muslim, she would be on her own for the next 3 years, pursuing her doctorate in computer science at the University of Cambridge in the United Kingdom. In *Girl Decoded*, el Kaliouby and coauthor Carol Colman have created a riveting memoir of a "nice Egyptian girl" who, despite cultural conditioning that encouraged her to put her duties as a wife and mother first, went on to pursue her professional dreams. She would become a pioneer in the emerging field of artificial emotional intelligence (emotion AI), where researchers seek to build computers that can sense and respond to human emotions.

El Kaliouby was inspired by the influential 1997 book *Affective Computing* by Rosalind Picard. Drawing on findings from neuroscience, Picard's central argument was that if we want smarter computers that interact more naturally with us, we must endow machines with the ability to recognize, understand, and even express emotions. This idea ran counter to conventional wisdom, which posited that pure logic was the highest form of AI.

In her first presentation at Cambridge, el Kaliouby announced that she planned to teach computers to read facial expressions, a proxy for a person's mental state. At the time, computers could barely tell apart a human face and a piece of fruit. The project was audacious.

A fellow student remarked that his brother, who was autistic, had great trouble understanding facial expressions. (Many people on the autism spectrum struggle to interpret nonverbal cues that indicate a person's mood or emotions.) Perhaps her work could ultimately help autistic people, he suggested.

El Kaliouby was intrigued and approached Simon Baron-Cohen, a prominent autism researcher on campus, to learn more. Baron-Cohen, it turned out, was building an exhaustive catalog of facial expressions to serve as a guide for those on the spectrum, and he allowed el Kaliouby to use the database to train her algorithm. Her face-reading code, "the Mind Reader," became a reality in 2004. "If I could have done an Egyptian *zaghroota*, the ululation of joy my people make at weddings, I would have," she confides.



Rana el Kaliouby demonstrates Affectiva's facial expression recognition technology in 2018.

El Kaliouby writes with candor about the challenges she faced during her training. She was homesick, and Cambridge's cold weather added to her misery. "I was lonely and my work was just creeping along," she admits. She barely saw her husband, who had founded a software company in Egypt, and their friends joked that she got pregnant with the couple's first child "over the Internet."

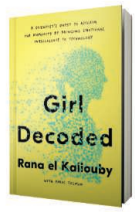
In 2004, el Kaliouby had the chance to show her work to the woman who had originally inspired her to study emotion AI. Rosalind Picard, who was visiting the Cambridge Computer Laboratory, was extremely impressed with el Kaliouby's work and offered her a postdoc position in the Affective Computing group at the Massachusetts Institute of Technology in Boston, Massachusetts.

El Kaliouby's reminiscences offer glimpses into her background and cultural upbringing. We learn, for example, that Arab men are conventionally referred to as the fathers of their eldest sons; for instance, "Abu Mohammed" means "father of Mohammed." If a man has no sons, then he is typically called by his own first name. El Kaliouby recounts how, once, when she was visiting her father's office in Abu Dhabi, a male colleague referred to him as "Abu Rana." She realized that her father must have spoken of her at work with pride, and by using this naming convention, his colleague was acknowledging the young woman's achievements. "I was deeply touched by the gesture," she writes.

In 2009, el Kaliouby and Picard founded the startup Affectiva to bring the emotion-sensing technology they had developed together in the laboratory to the marketplace. Many industries now put the company's software to myriad uses: from social robotics and market research to monitoring systems that check for distraction or drowsiness in drivers. In 2018, el Kaliouby was named to *Fortune* magazine's "40 Under 40" list of the most influential young people in the world of business.

Girl Decoded is an affecting memoir that highlights the tension between one woman's upbringing and her aspirations. Such life stories, told well, have the power to uplift us all. ■

Girl Decoded: A Scientist's Quest to Reclaim Our Humanity by Bringing Emotional Intelligence to Technology
Rana el Kaliouby with Carol Colman
Currency, 2020. 352 pp.



The reviewer is a science journalist in Boston, MA 02144, USA. Email: v.vijaysree@gmail.com

10.1126/science.abc3555

PHOTOS: AP PHOTO/ELISE AMENDOLA

Published by AAAS

LETTERS

A new national park could give China's depleted Hainan gibbon (*Nomascus hainanus*) population a better chance of recovery.

Edited by **Jennifer Sills**

Recovery hopes for the world's rarest primate

China was home to the world's second-highest diversity of gibbons in the 20th century, but these species have all suffered extensive population declines, and regional species losses continue (*1*, *2*). The Hainan gibbon (*Nomascus hainanus*), formerly widespread across Hainan Island (*3*), is now the rarest primate and among the rarest mammals (*4*). It represents one of the oldest gibbon lineages (*5*) but has received limited conservation attention compared with many other species in China. Hunting and habitat loss have reduced the species to a single high-elevation population in Bawangling National Nature Reserve (*4*, *6*). Restricted to 15 km^2 of fragmented rain forest within this reserve, the Hainan gibbon has persisted with fewer than 30 individuals for several decades (*3*, *6*, *7*). With China's plan to establish a national park on Hainan, the species has a renewed chance at recovery, but challenges remain.

Although the Hainan gibbon population received inconsistent financial support for protective measures in the 1980s and 1990s (*6*), conservation efforts since then have yielded promising signs. Two newly formed social groups were discovered in 2015 and 2019, and based on these positive trends, the total population size now likely exceeds 30 individuals (*8*). However, the species has extremely low genetic diversity (*9*), and forest degradation and conversion present barriers to the range expansion necessary for recovery (*7*). Increased enforcement has eliminated hunting as a threat (*6*, *7*), but the tiny population is inherently vulnerable to stochastic events such as emerging zoonotic disease outbreaks and extreme climatic events.

In 2015, the National Park system was established as China's new protected area framework (*10*). Hainan Tropical Rainforest National Park will be established in 2020 (*11*) and will cover almost one-seventh of Hainan's land area. The Hainan gibbon, as Hainan's endemic primate (*2*), has been designated the national park's flagship species (*12*). Raising the species' profile is an important step, but crucial data gaps remain, and National Park management must be evidence-based and guided by new research. In addition to increasing funding, it is essential to improve monitoring with innovative technologies, to understand habitat structure and resource distribution, and to implement science-led forest restoration. In situ conservation remains the priority, but feasibility of intensive recovery strategies under potential emergency scenarios must be appraised (*7*). Assessment of future climate change impacts is also urgently needed. Effective Hainan gibbon conservation and long-term protection of China's tropical biodiversity require robust scientific understanding, including insights from the recovery of other highly threatened species.

Hui Liu[1]*, Heidi Ma[2], Susan M. Cheyne[3], Samuel T. Turvey[2]

[1]Wuzhishan National Long Term Forest Ecosystem Research Station, College of Forestry, Hainan University, Haikou 570228, China. [2]Institute of Zoology, Zoological Society of London, London NW1 4RY, UK. [3]Faculty of Social Sciences, Oxford Brookes University, Oxford OX3 0BP, UK.
*Corresponding author.
Email: liuhui@hainanu.edu.cn

REFERENCES AND NOTES

1. P. Fan, *Biol. Conserv.* **210**, 29 (2017).
2. B. Li *et al.*, *Biodivers. Conserv.* **27**, 3301 (2018).
3. Z. Liu, S. Yu, X. Yuan, *Chin. Wildl.* **6**, 1 (1984).
4. T. Geissmann, W. Bleisch, *Nomascus hainanus* (The IUCN Red List of Threatened Species, 2008).
5. V. N. Thinh *et al.*, *BMC Evol. Biol.* **10**, 74 (2010).
6. B. P. L. Chan, J. R. Fellowes, T. Geissmann, J. Zhang, "Hainan gibbon status survey and conservation action plan" (Kadoorie Farm & Botanic Garden, 2005).
7. S. T. Turvey *et al.*, "International conservation planning workshop for the Hainan gibbon: Final report" (Zoological Society of London/IUCN SSC Conservation Breeding Specialist Group, 2015).
8. J. V. Bryant *et al.*, *Int. J. Primatol.* **37**, 534 (2016).
9. J. V. Bryant *et al.*, *Mol. Ecol.* **25**, 3540 (2016).
10. J. Li *et al.*, *Science* **351**, 1160 (2016).
11. "China makes headway in protecting tropical rainforests," *Xinhua* (2020).
12. "Speed up the construction of Hainan Rainforest National Park in 2020," *ChinaNews* (2020). http://www.hi.chinanews.com/hnnew/2020-01-23/513006.html [in Chinese].

COMPETING INTERESTS

S.M.C. is vice-chair of the International Union for Conservation of Nature/Species Survival Commission Primate Specialist Group's Section on Small Apes.

10.1126/science.abc1402

Cetaceans under threat in South China Sea

According to historical whaling and stranding records (*1*), the South China Sea is home to more than one-third of extant cetacean species on Earth, all of which are listed in Appendix I or II of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) (*2*). Recent data gleaned from the

PHOTO: YANG GUANYU/XINHUA/ALAMY STOCK PHOTO

Published by AAAS

ecological knowledge of local fishers (*3*) and field surveys (*4*) add to the evidence that part or all of the South China Sea is an important cetacean area with high species diversity that deserves special conservation attention but has been previously overlooked (*5*). Overfishing, together with other anthropogenic activities in the South China Sea, is putting cetaceans in grave danger.

The South China Sea contains lucrative fisheries (*6*, *7*). However, in recent decades, illegal, unreported, and unregulated fishing has led to region-wide overfishing (*6*, *7*), posing a major threat to cetaceans (*8*). Overfishing limits cetaceans' nutrition through prey depletion (*9*) and is closely tied to vessel strikes and bycatch of hundreds of thousands of cetaceans (*10*, *11*). Many fishing vessels also produce underwater noises that damage cetaceans' hearing and disturb their behavior, causing disorientation (*12*). As a result, cetaceans could disappear from the South China Sea if no conservation actions are taken.

Cetaceans are important megafauna and valuable marine biological resources; saving them is vital to maintaining marine biodiversity and ecosystems. Communication with the wider science, conservation, and policy communities about the need to conserve cetaceans in the South China Sea is the first step toward raising awareness and inciting policy changes by regional governments to manage and regulate fisheries and other anthropogenic activities. Meanwhile, regional governments should invest in research to investigate and identify important cetacean areas in the South China Sea, independent of any political or socioeconomic concerns. This research can inform the design and management of marine protected areas, where guidelines or regulations are needed to control anthropogenic activities, including fishing. Finally, the migratory nature of both fishes and cetaceans requires all South China Sea claimant nations to realize that they have a legal obligation to cooperate on fisheries management. National plans of action against illegal, unreported, and unregulated fishing need to be harmonized throughout the region, which will both safeguard the food security and livelihoods of the communities in and around the South China Sea and protect vulnerable cetaceans from continued harm.

Songhai Li[1,2]*, Mingli Lin[1], Francesco Caruso[1], Lijun Dong[1], Wenzhi Lin[1], Massimiliano Rosso[1,3], Alessandro Bocconcelli[1,4]
[1]Marine Mammal and Marine Bioacoustics Laboratory, Institute of Deep-sea Science and Engineering, Chinese Academy of Sciences, Sanya, 572000, China. [2]Tropical Marine Science Institute, National University of Singapore, 119227, Singapore. [3]CIMA Research Foundation, Savona, 17100, Italy. [4]Advanced Ocean Physics and Engineering Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA.
*Corresponding author. Email: lish@idsse.ac.cn

REFERENCES AND NOTES

1. P. Wang, Ed., *Chinese Cetaceans* (Chemical Industry Press, Beijing, 2011) [in Chinese].
2. CITES, Appendices I, II, and III (2019); https://cites.org/eng/app/appendices.php.
3. M. Lin *et al.*, *Ocean Coast. Manage.* **172**, 117 (2019).
4. "Deep-diving cetacean survey in the South China Sea: Disclosing seven cetaceans including sperm whales," *Beijing News* (2019); http://news.sina.com.cn/c/2019-05-11/doc-ihvhiews1206140.shtml [in Chinese].
5. "Important Marine Mammal Areas: 46 new candidates," IUCN (2018); www.iucn.org/news/world-commission-protected-areas/201803/important-marine-mammal-areas-46-new-candidates.
6. S. K. Chang *et al.*, *Mar. Pol.* **116**, 103716 (2020).
7. J. Li, R. Amer, *J. Int. Wildl. Law Pol.* **18**, 139 (2015).
8. M. Liu *et al.*, *Anim. Conserv.* **20**, 182 (2017).
9. G. Bearzi *et al.*, *Biol. Conserv.* **127**, 373 (2006).
10. F. Ritter, S. Panigada, in *World Seas: An Environmental Evaluation*, C. Sheppard. Ed. (Elsevier, ed. 2, 2019), pp. 531–547.
11. J. J. Alava *et al.*, *Mar. Pol.* **110**, 102769 (2019).
12. T. A. Jefferson, M. A. Webber, R. L. Pitman, *Marine Mammals of the World: A Comprehensive Guide to their Identification* (Academic Press, 2008).

10.1126/science.abc7557

Tenure and promotion after the pandemic

To compensate for declines in productivity induced by the pandemic, many universities have automatically extended tenure clocks by 1 year (*1*). This move is necessary but not sufficient. Tenure clock extensions disadvantage some groups. For example, in economics, women on longer clocks due to parental leave get tenure at lower rates than men (*2*). Many men use leave to produce articles, whereas women are more likely to care for children. Tenure committees often fail to account for differences in how leave time is spent. The coronavirus disease 2019 (COVID-19) pandemic will produce additional inequalities. Stay-at-home and public safety orders preclude some research, such as human subjects and laboratory work, while creating opportunities for others. Now is the time to develop strategies to mitigate the inequitable effects of the quarantine.

External reviewers and tenure and promotion committees should make three adjustments to the evaluation process. First, they should instruct candidates to pick the 6 best years of their record and require that evaluators and committees assess only the quality and impact of research, teaching, and service from those years, not the total years after completing their Ph.D. (or after attaining tenure, in the case of full professors). Second, evaluators should require a COVID-19 impact statement that explains the research, teaching, and service that candidates were able or unable to do, infrastructural or financial constraints, and obligations including child and elder care. Third, reviewers and committees should consider qualitative and holistic assessments in addition to quantitative evaluations such as number of publications, citations, impact factors, research expenditures, and teaching scores. These indicators could potentially carry even more biases after COVID-19 (*3*–*6*). Let's treat the pandemic as an opportunity to adopt new standards for advancement that are fair to caregivers and people with diverse research agendas.

Mala Htun
Department of Political Science, University of New Mexico, Albuquerque, NM 87131, USA. Email: malahtun@gmail.com

REFERENCES AND NOTES

1. L. Scungio, "Adapting in times of crisis: Navigating tenure clock stoppage," *Collaborative on Academic Careers in Higher Education* (blog) (2020); https://coache.gse.harvard.edu/blog/adapting-times-crisis-navigating-tenure-clock-stoppage.
2. H. Antecol, K. Bedard, J. Stearns, *Am. Econ. Rev.* **108**, 2420 (2018).
3. M. L. Dion, J. L. Sumner, S. McLaughlin Mitchell, *Polit. Anal.* **26**, 312 (2018).
4. D. L. Teele, K. Thelen, *PS: Polit. Sci. Polit.* **50**, 433 (2017).
5. F. Mengel, J. Sauermann, U. Zölitz, *J. Eur. Econ. Assoc.* **17**, 535 (2019).
6. F. R. Jensenius *et al.*, *PS Polit. Sci. Polit.* **51**, 820 (2018).

10.1126/science.abc7469

TECHNICAL COMMENT ABSTRACTS

Comment on "RNA-guided DNA insertion with CRISPR-associated transposases"

Phoebe A. Rice, Nancy L. Craig, Fred Dyda

Strecker *et al.* (Research Article, 5 July 2019, p. 48) described a system for exploiting a Tn7-type transposon-encoded CRISPR-Cas system to make RNA-guided, programmable insertions. Although this system has great promise, we note that the well-established biochemistry of Tn7 suggests that the particular system used may insert not only the transposon but also the entire donor plasmid.

Full text: dx.doi.org/10.1126/science.abb2022

Response to Comment on "RNA-guided DNA insertion with CRISPR-associated transposases"

Jonathan Strecker, Alim Ladha, Kira S. Makarova, Eugene V. Koonin, Feng Zhang

Rice *et al.* suggest that the CRISPR-associated transposase ShCAST system could lead to additional insertion products beyond simple integration of the donor. We clarify the outcomes of ShCAST-mediated insertions in *Escherichia coli*, which consist of both simple insertions and integration of the donor plasmid. This latter outcome can be avoided by use of a 5′ nicked DNA donor.

Full text: dx.doi.org/10.1126/science.abb2920

Published by AAAS

Cite as: P. A. Rice *et al.*, *Science* 10.1126/science.abb2022 (2020).

Comment on "RNA-guided DNA insertion with CRISPR-associated transposases"

Phoebe A. Rice[1]*, Nancy L. Craig[2], Fred Dyda[3]

[1]Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL, USA. [2]Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [3]Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, USA.

*Corresponding author. Email: price@uchicago.edu

Strecker *et al.* (Research Articles, 5 July 2019, p. 48) described a system for exploiting a Tn7-type transposon-encoded CRISPR-Cas system to make RNA-guided, programmable insertions. Although this system has great promise, we note that the well-established biochemistry of Tn7 suggests that the particular system used may insert not only the transposon but also the entire donor plasmid.

The *Scytonema hofmanni* CRISPR-associated transposase (ShCAST) system described by Strecker *et al.* (*1*) is based on a Tn7-like transposon that carries its own nuclease-defective CRISPR-Cas system, which it uses for RNA-guided target choice (*2*). The canonical *Escherichia coli* transposon Tn7 uses two enzymes to create double-strand breaks at the donor-transposon junctions: TnsB to cleave one strand (and subsequently attach it to target DNA) and TnsA to cleave the other (second) DNA strand, leading to excision of the transposon from a donor plasmid followed by its insertion into a target plasmid. When TnsA is mutated, however, the products are not these simple insertions of the transposon, but rather co-integrates in which the entire donor plasmid, flanked by copies of the transposon, is inserted into the target (*3*). Similar co-integrate products are also created by systems such as Mu transposase, which is a TnsB homolog that lacks a TnsA partner (*4*).

The ShCAST Tn7-like transposon used by Strecker *et al.* is naturally lacking in TnsA. Although some DDE transposases harbor insertion domains that trigger other mechanisms for cleavage of the second strand, ShCAST TnsB does not. Homologies to the other systems noted above therefore suggest that their products may not be solely simple insertions as cartooned, but instead may include fusions of the donor plasmid and target. Unfortunately, no experiments were reported that would distinguish between the two possibilities, such as measuring the size of purified "pInsert" product plasmids containing the transposon or testing for the presence of donor DNA in the final products. Although the ShCAST integration system is suggested to bypass the need for homologous recombination–mediated DNA repair normally found in some CRISPR/Cas applications, processing of co-integrate products to remove the donor DNA would likely still require these recombination functions unless another nuclease processed the second strand at the donor site.

REFERENCES

1. J. Strecker, A. Ladha, Z. Gardner, J. L. Schmid-Burgk, K. S. Makarova, E. V. Koonin, F. Zhang, RNA-guided DNA insertion with CRISPR-associated transposases. *Science* **365**, 48–53 (2019). doi:10.1126/science.aax9181 Medline
2. J. E. Peters, K. S. Makarova, S. Shmakov, E. V. Koonin, Recruitment of CRISPR-Cas systems by Tn7-like transposons. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E7358–E7366 (2017). doi:10.1073/pnas.1709035114 Medline
3. E. W. May, N. L. Craig, Switching from cut-and-paste to replicative Tn7 transposition. *Science* **272**, 401–404 (1996). doi:10.1126/science.272.5260.401 Medline
4. A. B. Hickman, F. Dyda, DNA Transposition at Work. *Chem. Rev.* **116**, 12758–12784 (2016). doi:10.1021/acs.chemrev.6b00003 Medline

6 February 2020; accepted 11 May 2020
Published online 5 June 2020
10.1126/science.abb2022

Cite as: J. Strecker *et al.*, *Science* 10.1126/science.abb2920 (2020).

Response to Comment on "RNA-guided DNA insertion with CRISPR-associated transposases"

Jonathan Strecker[1,2,3,4], Alim Ladha[1,2,3,4], Kira S. Makarova[5], Eugene V. Koonin[5], Feng Zhang[1,2,3,4]*

[1]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. [2]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [3]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [4]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [5]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA.

*Corresponding author. Email: zhang@broadinstitute.org

Rice *et al.* suggest that the CRISPR-associated transposase ShCAST system could lead to additional insertion products beyond simple integration of the donor. We clarify the outcomes of ShCAST-mediated insertions in *Escherichia coli*, which consist of both simple insertions and integration of the donor plasmid. This latter outcome can be avoided by use of a 5′ nicked DNA donor.

Rice *et al.* (*1*) raise an important point about the mechanism of integration of the type V CRISPR-associated transposase (CAST) systems we recently characterized (*2*). Although CAST loci belong to the Tn7-like class of transposons, additional analysis indicates that type V CAST loci are most closely related to transposons of the Tn*5053* family (*3*, *4*), a distinct family of transposons that has not yet been mechanistically characterized. In particular, the Tn*5053* family transposons lack TnsA, the enzyme responsible for the 5′ donor cleavage in the Tn7 transposon (*5*, *6*), and result in a co-integrate product that is resolved by the site-specific recombinase TniR (*4*). Thus, in CAST systems, the 5′ donor ends might not be cleaved, resulting in a co-integrate product containing duplicated cargo DNA and the donor backbone. Alternatively, a CAST protein might either cleave the 5′ donor end or help resolve the co-integrate to yield a simple insertion.

To investigate the exact structure of the insertion product, we performed nanopore sequencing of 15 *Scytonema hofmanni* CAST (ShCAST)–mediated genome insertions in *E. coli* and found nine simple insertions and nine co-integrates across several target sites (Fig. 1A). Similarly, a genetic assay using a plasmid target revealed about 20% co-integrate insertions (Fig. 1B). A tentative model is that the initial insertion product is a co-integrate that can be resolved by the cellular DNA recombination and repair systems. However, we cannot rule out that CAST components contribute to 5′ donor cleavage or co-integrate resolution in *E. coli* or the native host. Notably, all instances of CAST insertion in cyanobacterial genomes identified to date are simple insertions. We will continue to investigate these as a part of our ongoing studies.

Use of linear or 5′ nicked DNA donors prevents co-integrate formation (Fig. 1C), thereby providing an approach for applying CAST for homologous recombination–independent genome engineering. Continued investigation of the mechanism of CAST is expected to yield a deeper understanding of the biology of this remarkable DNA integration system and propel its development as a molecular technology.

REFERENCES

1. P. A. Rice, N. L. Craig, F. Dyda, Comment on "RNA-guided DNA insertion with CRISPR-associated transposases". *Science* **368**, eabb2022 (2020).
2. J. Strecker, A. Ladha, Z. Gardner, J. L. Schmid-Burgk, K. S. Makarova, E. V. Koonin, F. Zhang, RNA-guided DNA insertion with CRISPR-associated transposases. *Science* **365**, 48–53 (2019). doi:10.1126/science.aax9181 Medline
3. G. Y. Kholodii, O. V. Yurieva, O. L. Lomovskaya, Z. Gorlenko, S. Z. Mindlin, V. G. Nikiforov, Tn5053, a mercury resistance transposon with integron's ends. *J. Mol. Biol.* **230**, 1103–1107 (1993). doi:10.1006/jmbi.1993.1228 Medline
4. G. Y. Kholodii, S. Z. Mindlin, I. A. Bass, O. V. Yurieva, S. V. Minakhina, V. G. Nikiforov, Four genes, two ends, and a res region are involved in transposition of Tn5053: A paradigm for a novel family of transposons carrying either a mer operon or an integron. *Mol. Microbiol.* **17**, 1189–1200 (1995). doi:10.1111/j.1365-2958.1995.mmi_17061189.x Medline
5. M. C. Biery, M. Lopata, N. L. Craig, A minimal system for Tn7 transposition: The transposon-encoded proteins TnsA and TnsB can execute DNA breakage and joining reactions that generate circularized Tn7 species. *J. Mol. Biol.* **297**, 25–37 (2000). doi:10.1006/jmbi.2000.3558 Medline
6. R. J. Sarnovsky, E. W. May, N. L. Craig, The Tn7 transposase is a heteromeric complex in which DNA breakage and joining activities are distributed between different gene products. *EMBO J.* **15**, 6348–6361 (1996). doi:10.1002/j.1460-2075.1996.tb01024.x Medline

24 March 2020; accepted 11 May 2020
Published online 5 June 2020
10.1126/science.abb2920

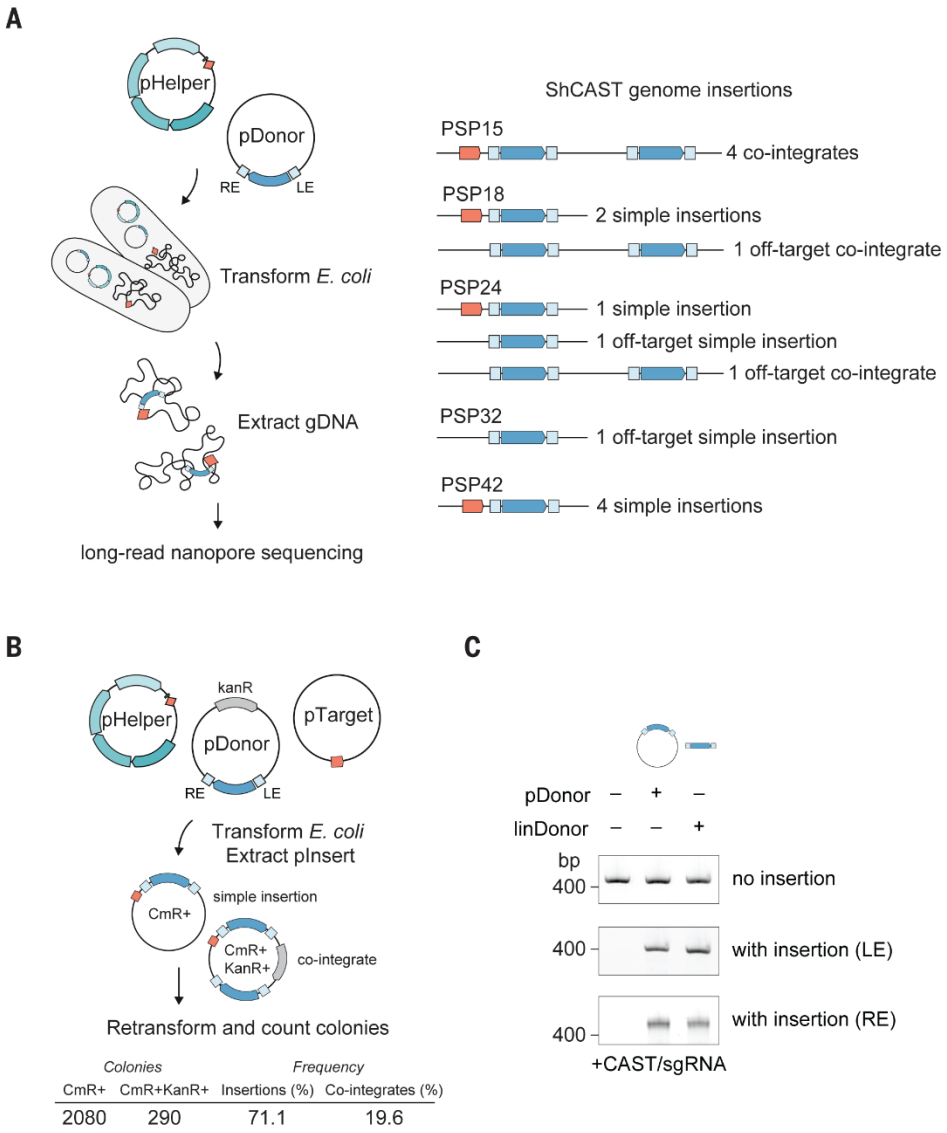


Fig. 1. Characterization of CAST insertion products. (**A**) Schematic of genome targeting experiment and summary of nanopore sequencing results. (**B**) Genetic assay for plasmid targeting. pInserts were retransformed and selected on CmR+ and CmR+KanR+ plates to determine the fraction of co-integrate insertions. The total insertion frequency was determined by droplet digital PCR and used to calculate the co-integrate rate. (**C**) In vitro reactions with purified CAST proteins using plasmid donor or PCR-amplified linear donor.

RESEARCH

IN *SCIENCE* JOURNALS

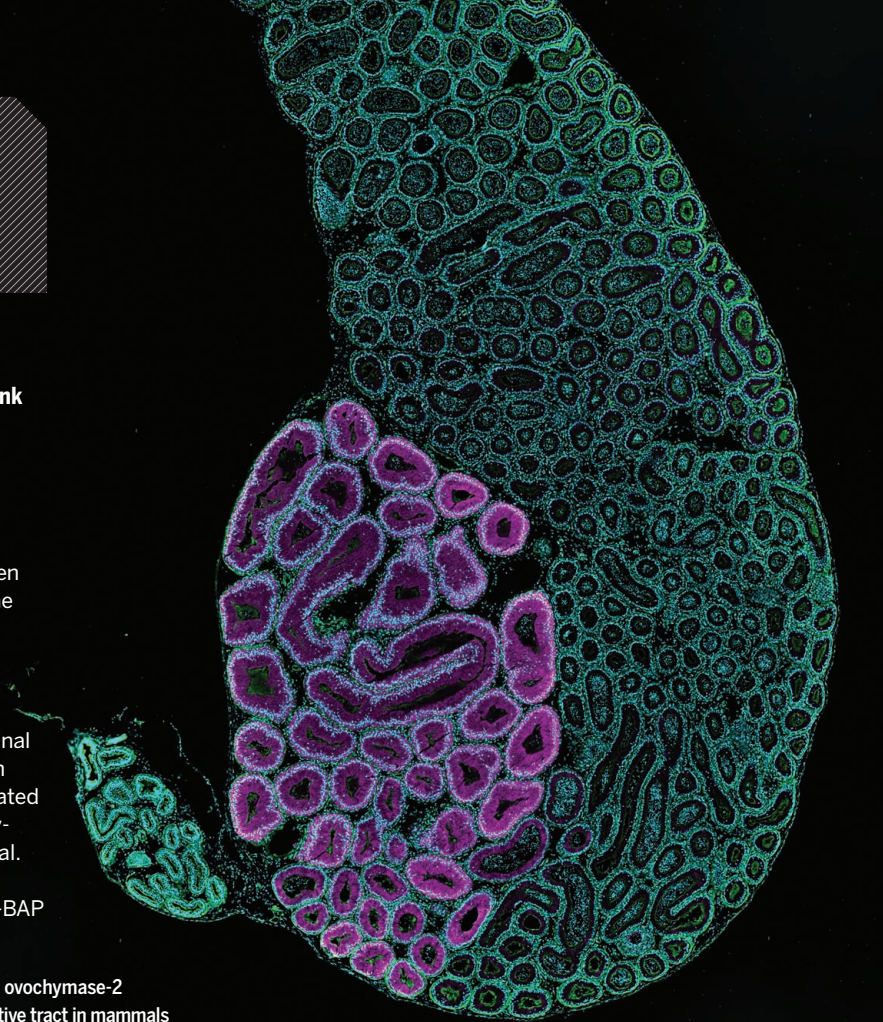
Edited by **Michael Funk**

REPRODUCTIVE BIOLOGY

Local control of sperm maturation

Newly produced spermatozoa within the testis do not have fertilizing ability but become fully functional when they mature in the epididymis. The development of the epididymis itself is dependent on testicular factors arriving via luminal flow. Improper signaling between the testis and epididymis is hypothesized to result in male infertility. Kiyozumi *et al*. identified NELL2 as a testicular luminal protein that binds to its receptor, ROS1, on the luminal epididymis surface and induces epididymal differentiation (see the Perspective by Lord and Oatley). In turn, differentiated epididymis secretes a fertility-essential protease, ovochymase-2, to make spermatozoa fully mature and functional. Thus, testis-epididymis interorgan communication by this "lumicrine" regulation ensures mammalian reproduction. —BAP

Science, this issue p. 1132; see also p. 1053



Fluorescence microscopy image revealing localization of the protease ovochymase-2 (purple) within the caput epididimus, which is part of the male reproductive tract in mammals

CREDIT: KIYOZUMI *ET AL.*

WEARABLE DEVICES

Improving ionic thermoelectrics

Using ions as charge carriers in thermoelectric devices usually requires using either thermal diffusion or redox reactions at two electrodes with different temperatures. Han *et al*. leveraged both of these strategies to develop a gelatin-based ionic thermoelectric device that uses alkali salts and an iron-based redox couple to generate a large thermopower. This device is capable of generating useful amounts of energy from body heat. —BG

Science, this issue p. 1091

PLASTIC POLLUTION

Not just settling

What controls the distribution of microplastics on the deep seafloor? Kane *et al*. show that the answer to that question is more complicated than particles simply settling from where they are found on the sea surface (see the Perspective by Mohrig). Using data that they collected off the coast of Corsica, the authors show that thermohaline-driven currents can control the distribution of microplastics by creating hotspots of accumulation, analogous to their role in causing focused areas of seafloor sediment deposition. Such currents also supply oxygen and nutrients to deep-sea benthos, so deep-sea biodiversity hotspots are also likely to be microplastic hotspots. —HJS

Science, this issue p. 1140; see also p. 1055

CORONAVIRUS

Smothering fecal-oral coronavirus spread

Diarrhea is a common symptom in patients with coronavirus disease 2019 (COVID-19). Zang *et al*. used organoid cultures of cells from the epithelial lining of human small and large intestine as an in vitro model system with which to study viral entry and replication in enterocytes. Mature enterocytes expressing the viral receptor were susceptible to productive infection, which was also stimulated by the expression of a protease involved in viral entry. A subset of patients with COVID-19 shed high amounts of viral RNA in feces, but experiments with simulated human colonic fluid suggested that any shed virus would be rapidly inactivated during transit through the colon. —IRW

Sci. Immunol. **5**, eabc3582 (2020).

TOPOLOGICAL OPTICS

Topological insulators in the spotlight

In addition to having an insulating interior while at the same time supporting conducting surface states, topological insulators have many other interesting properties. Higher-order topological insulating states, where regions of interest are along edges and at corners, have been difficult to identify unambiguously. Peterson *et al*. developed a theoretical framework to help identify and characterize these exotic states, including a new topological marker—the fractional charge density—that can be used to detect topological states of matter when the spectroscopic probe of gapless surface states is not accessible. The agreement between experimental work and theory is encouraging for applicability to other topological platforms. —ISO

Science, this issue p. 1114

CANCER

How cancer cells adapt to stress

Bacteria adapt to harsh conditions such as antibiotic

Published by AAAS

exposure by acquiring new mutations, a process called stress-induced mutagenesis. Cipponi *et al*. investigated whether similar programs of mutagenesis play a role in the response of cancer cells to targeted therapies. Using in vitro models of intense drug selection and genome-wide functional screens, the authors found evidence for an analogous process in cancer and showed that it is regulated by the mammalian target of rapamycin (mTOR) signaling pathway. This pathway appears to mediate a stress-related switch to error-prone DNA repair, resulting in the generation of mutations that facilitate the emergence of drug resistance. —PAK

Science, this issue p. 1127

ARCHAEOLOGY

Timing the rise of maize in Mesoamerica

Many lines of evidence suggest that maize (*Zea mays*) became a dietary staple across ancient Mesoamerica. However, there has been little direct evidence of its consumption, and the timing of how it came to dominate the diet of the peoples of the region is unknown. Using stable isotopic evidence from human skeletons excavated from two rock shelter sites in Belize, Kennett *et al*. show that there is no clear evidence of maize consumption by the sites' inhabitants before 4700 years ago. However, isotopes from more recent individuals show the increasing importance of maize in the diet, such that by 4000 years ago, maize had become a persistent dietary staple. —MSA

Sci. Adv. 10.1126/sciadv.aba3245 (2020).

ANALYTICAL CHEMISTRY

Perfluorocarbons' path into soils

Covering carbon chains with fluorines has produced a variety of useful nonstick coatings. However, growing concern about the toxicity and extraordinary environmental persistence of the underlying compounds is spurring a search for alternatives. The precise structure of these next-generation alternatives often remains a trade secret. Washington *et al*. sampled soils in New Jersey and then used mass spectrometry to assign plausible structures—incorporating chlorine and ether segments into the CF_2 chain—to compounds that appear to have emanated from their manufacture (see the Policy Forum by Gold and Wagner). The data can inform in-depth studies of these compounds' environmental transport and persistence. —JSY

Science, this issue p. 1103; see also p. 1066

PHOTOS (FROM LEFT): DBIMAGES/ALAMY STOCK PHOTO; LILLY TROTT/SHUTTERSTOCK

Maize, a global staple crop seen here growing in a field in El Salvador, rose to dominate diets in Mesoamerica by 4000 years before present.

IN OTHER JOURNALS

Edited by **Caroline Ash** *and* **Jesse Smith**

MATERIALS SCIENCE

Better food preservation

Fresh fruits and vegetables are a key part of a healthy diet but there can be considerable waste caused by spoilage after harvest. Shelf life can be extended by preventing microbial growth, dehydration, or storage at higher temperatures, but adding a wax coating, creating a reduced oxygen environment, or using refrigeration can be expensive and/or time consuming and may alter the taste of the food. Jung *et al*. developed an egg-sourced albumin coating reinforced with nanocrystalline cellulose that can be made from waste materials. When coated onto banana, avocado, papaya, and strawberry, shelf life was extended by a week, with reduced external browning and internal ripening. The coatings are safe to ingest but are also easily removed through washing. —MSL

Adv. Mater. 10.1002/adma.201908291 (2020).

Postharvest spoilage of fresh fruits and vegetables is reduced by a safe and easily removed coating made with egg albumin.

DEVELOPMENT

Making cartilage throughout life

The skeletons of newborn mammals are soft and pliable because they are composed primarily of cartilage. During growth to adulthood, most cartilage is replaced by bone. The remaining cartilage, such as that found in the joints, does not readily regenerate, so joints deteriorate with age. By contrast, elasmobranch fish make cartilage throughout their

Published by AAAS

FOOD SECURITY

Scarcity in times of abundance

A cornerstone of food security is stability of access to the food supply. In a pioneering study of fishing communities in Amazonian seasonally flooded forest, Tregidgo *et al*. show how seasonal fluctuations in access can lead to scarcity even when the resource is abundant. Fishing becomes more challenging in the high-water season, leading to severe food insecurity for the more deprived households in the community. Communities such as these that depend on a wildlife resource for a substantial proportion of their diet may become further disadvantaged as seasonal fluctuations in access to food become further exacerbated by climate change. —AMS

People Nat. 10.1002/pan3.10086 (2020).

For communities living along the Amazon, greater seasonal fluctuations in water levels brought by climate change are likely to make access to fish resources more challenging.

lives. Marconi *et al*. studied the cartilaginous fish *Leucoraja erinacea*, or little skate, from embryo to adult and observed that progenitor cells that surround the cartilage skeleton are also present in the adult. These cells are induced to transform into chondrocytes after injury. Understanding cartilage repair in skates may offer inspiration for research into human joint repair. —BAP

eLife **9**, e53414 (2020).

PARASITOLOGY

How helminths trump diabetes

The prevalence of type 2 diabetes mellitus (T2DM) is inversely correlated with helminth infections in Asia. This may be because helminths have an immunomodulatory effect and thus dampen the type 1 (allergic type) immune responses underlying the proinflammatory state of T2DM. To test this idea, Rajamanickam *et al*. measured plasma levels of cytokines in 60 individuals living in rural India infected with *Strongyloides stercoralis*, a persistent nematode gut parasite, which in most people is symptomless. They compared these results with plasma cytokine concentrations from 58 T2DM patients showing no worm infection. Parasitized individuals showed significantly increased levels of interleukin-1 receptor alpha (IL-1Rα), which is typical of type 2 immune responses, and low levels of a wide range of proinflammatory cytokines and chemokines. When diabetic subjects were treated for parasites, their proinflammatory state, typically marked by increased IL-1β and IL-6, partially rebounded. —CA

PLOS Neg. Trop. Dis. **14**, e0008101 (2020).

IMMUNOLOGY

Neutrophils avoid a traffic jam

Neutrophils are the most abundant immune cell in the circulation and are typically the first responders to sites of infection or injury. How large numbers of neutrophils can efficiently travel through capillary networks is a mystery. Wang *et al*. investigated neutrophil trafficking in mouse liver using intravital microscopy and found that groups of neutrophils diverged at capillary bifurcations by traveling in an alternating pattern. This phenomenon was then studied in a controlled fashion using microfluidic chips connected to a chemoattractant chamber. Neutrophils were able to bias the decisions made by their companions at bifurcations by altering both hydraulic resistance and chemoattractant gradients. It is likely that similar mechanisms are widely used to coordinate complex immune responses. —STS

Nat. Commun. **11**, 2385 (2020).

EXOPLANET ATMOSPHERES

Different as night and day

The atmospheric compositions of exoplanets are usually determined by the technique of transit spectroscopy: When the planet transits between its host star and Earth, starlight passes through its atmosphere close to local dawn and dusk, imprinting additional absorption lines on the stellar spectrum. Pluriel *et al*. considered how the results are affected if the star heats the planet enough for atmospheric molecules to be dissociated on the planet's day side and recombine on the night side. They found that transit spectroscopy then produces highly biased atmospheric compositions differing from the true values by up to three orders of magnitude. More sophisticated methods are needed to determine the atmospheric compositions of hot exoplanets. —KTS

Astron. Astrophys. **636**, A66 (2020).

SCIENTIFIC WORKFORCE

To a postdoc and beyond

To be a postdoctoral fellow today is to exist between a short-term research position and the reality of the academic job market. Nowell *et al*. sought to identify professional development (PD) opportunities used by postdocs to navigate a changing career landscape. Using quantitative and qualitative data collected from more than 500 current postdocs, the authors show that although postdocs engaged in diverse PD opportunities, these strategies were not always useful or career enhancing. By describing postdoc perceptions of PD, in parallel with their perceived usefulness, this study provides a comprehensive understanding of the misalignment of postdoc needs with available opportunities and can serve as a manual for helping academic institutions direct valuable resources toward the most effective postdoc PD strategies. —MMc

Palgrave Commun. **6**, 95 (2020).

PHOTO: ROBERTHARDING/ALAMY STOCK PHOTO

Published by AAAS

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

CORONAVIRUS

Applications of antibody testing

There has been a strong focus on testing to determine whether an individual has an active severe acute respiratory syndrome–coronavirus 2 (SARS-CoV-2) infection, but how do we find out if someone has been infected in the past? In a Perspective, Krammer and Simon discuss the key applications of serological tests to detect antibodies against SARS-CoV-2. Serological testing could answer key questions about whether neutralizing antibodies form, and if they do, serological studies could assess the duration of immune protection. Such tests could also identify recovered patients with high amounts of antibody who could donate blood serum for therapeutic use. In addition, population serosurveys could inform mitigation practices, which could be especially important during subsequent predicted waves of infection. However, it is important to ensure that tests are accurate and sufficiently reliable first. —GKA

Science, this issue p. 1060

COMPUTER SCIENCE

From bottom to top

The doubling of the number of transistors on a chip every 2 years, a seemly inevitable trend that has been called Moore's law, has contributed immensely to improvements in computer performance. However, silicon-based transistors cannot get much smaller than they are today, and other approaches should be explored to keep performance growing. Leiserson *et al*. review recent examples and argue that the most promising place to look is at the top of the computing stack, where improvements in software, algorithms, and hardware architecture can bring the much-needed boost. —JS

Science, this issue p. 1079

PLASMID EVOLUTION

Agrobacteria virulence writ large

Plasmids are widespread among bacteria and are important because they spread virulence and antibiotic resistance traits, among others. They are horizontally transferred between strains and species, so it is difficult to work out their evolution and epidemiology. Agrobacteria, a diverse grouping of species that infect plants, inject oncogenic Ti and Ri plasmids, which cause crown galls and hairy root diseases, respectively. The upside is that these plasmids have become valuable biotechnological tools. Weisberg *et al*. combed through an 80-year-old collection of *Agrobacterium* strains but found a surprisingly low diversity of plasmids. It is puzzling how limited the number of plasmid lineages is despite reported high levels of plasmid recombination, but what is clear is how plant production systems have influenced plasmid spread into various genomic backbones. —CA

Science, this issue p. 1080

STRUCTURAL BIOLOGY

Architecture of DNA-organizing complex

The highly conserved mammalian CTC1-STN1-TEN1 (CST) complex is critical for genome stability and telomere maintenance. Lim *et al*. solved the structure of the human CST complex using cryo–electron microscopy. CST forms an unprecedented and substantial decameric supercomplex triggered by telomeric single-stranded binding. This decameric form with single-stranded DNA–binding capacity of up to 10 telomeric repeats, suggested the possibility of CST organizing telomere overhangs into compact and restrictive structures in a manner similar to the nucleosome's organization of double-stranded DNA. This work provides a platform for understanding the mechanisms of various CST functions. —SYM

Science, this issue p. 1081

CHEMICAL PHYSICS

Watching electrons swarm ammonia

Liquid ammonia is unusual in its capacity to host electrons in stable solution, with vivid blue and bronze colors signifying the low- and high-concentration regimes, respectively. Buttersack *et al*. used photoelectron spectroscopy and accompanying theoretical simulations to track the precise energetic changes that ensued as steadily rising quantities of electrons were introduced by dissolved lithium, sodium, or potassium (see the Perspective by Isborn). The results point to a gradual transition from the dilute electrolyte solution of paired dielectrons to the more delocalized metallic structure at the highest concentrations. —JSY

Science, this issue p. 1086; see also p. 1056

ORGANIC CHEMISTRY

Hydrogenations that tolerate N–O bonds

Catalysts that add hydrogen to carbon-carbon, carbon-nitrogen, and carbon-oxygen double bonds are among the most widely used in synthetic chemistry. They are particularly adept at delivering just one of two mirror-image products. However, they may also target adjacent bonds in the compound that would be better left intact. Mas-Roselló *et al*. report that an iridium catalyst paired with a strong acid can hydrogenate C=N bonds without disturbing a weak N–O bond on the same nitrogen center. The reactions proceed at room temperature with high enantioselectivity. —JSY

Science, this issue p. 1098

NEUROSCIENCE

Making blind retinas see again

Photoreceptor degeneration is an important cause of blindness. Nelidova *et al*. used tunable, near-infrared sensors to render diseased photoreceptors light sensitive again (see the Perspective by Franke and Vlasits). Gold nanorods capable of detecting infrared light were coupled with an antibody to temperature-sensitive ion channels. When the nanorods absorbed light and converted it into heat, the coupled ion channels were gated by infrared light. In a mouse model of retinal degeneration, these ion channels were successfully targeted to cone photoreceptors, and responses to near infrared light could be detected. In the primary visual cortex, more cells responded to near-infrared stimuli in mice expressing these ion channels than in controls. By changing the length of the gold nanorods, the system could be tuned to different infrared wavelengths. —PRS

Science, this issue p. 1108; see also p. 1057

PALEOECOLOGY

Mangroves under sea level rise

The rate of sea level rise has doubled from 1.8 millimeters per year over the 20th century to ~3.4 millimeters per year in recent years. Saintilan *et al*. investigated the likely effects of this increasing rate of rise on coastal mangrove forest, a tropical ecosystem of key importance for coastal protection (see the Perspective by Lovelock). They reviewed data on mangrove accretion 10,000 to 7000 years before present, when the rate of sea level rise was even higher than today as a result of glacial ice melt. Their analysis suggests an upper threshold of 7 millimeters per year as the maximum rate of sea level rise

Published by AAAS

associated with mangrove vertical development, beyond which the ecosystem fails to keep up with the change. Under projected rates of sea level rise, they predict that a deficit between accretion and sea level rise is likely to commence in the next 30 years. —AMS

Science, this issue p. 1118; see also p. 1050

IMMUNOLOGY

Innate immune cells remember

Immunological memory is a phenomenon by which immune cells can quickly recognize an antigen that the host has previously encountered. Certain cells of the innate immune system exhibit memory-like responses know as trained immunity. Rapid, antigen-specific secondary (anamnestic) responses were long thought to be the domain of B and T cells. However, Dai *et al*. report that monocytes and macrophages can acquire memory specific for particular major histocompatibility complex I antigens using paired A-type immunoglobulin-like receptors (PIR-As) (see the Perspective by Dominguez-Andrés and Netea). This pathway contributes to recognition and rejection of allograft-transplanted tissue from a donor of the same species. Genetic depletion or blockade of PIR-As in mice diminished the rejection of kidney and heart allografts. This work, which expands immunological memory to include myeloid cells, points to targets that may improve organ transplantation outcomes in the future. —STS

Science, this issue p. 1122; see also p. 1052

MOLECULAR BIOLOGY

DNA barcodes in small packages

Under adverse environmental conditions, some microorganisms form spores that provide robust protection for genetic material. Qian *et al*. developed a system in which DNA barcodes are encapsulated inside nongerminating microbial spores and can be dispersed on objects or in the environment (see the Perspective by Nivala). These barcoded spores provide a durable, specific marker that can be read out quickly with simple equipment. When applied to soil, the spores can be transferred to and from objects around them, enabling tracking at meter-scale resolution. On plant leaves, the spores are not readily transferred, and the authors demonstrate a potential use for tracking agricultural products. —MAF

Science, this issue p. 1135; see also p. 1058

GENE EDITING

Enforced editing

Various autoimmune diseases could potentially be treated with regulatory T cells (T_{regs}), but there are many hurdles between this idea and clinical execution. Honaker *et al*. devised a gene-editing strategy to enforce the expression of FOXP3, the master T_{reg} transcription factor, in $CD4^+$ T cells isolated from human peripheral blood, thereby overcoming the limitations of T_{reg} isolation and expansion. The resulting stable FOXP3 expression enabled a suppressive phenotype in vitro, and the edited cells were also functional in both a xenogeneic graft-versus-host disease model and an experimental autoimmune encephalitis model. This approach has the potential to rapidly translate to clinical use. —LP

Sci. Transl. Med. **12**, eaay6422 (2020).

PHARMACOLOGY

More targeted endothelial protection

The bioactive lipid S1P exerts effects in diverse tissues. Unbiased agonists of the S1P receptor $S1P_1$ trigger a reduced lymphocyte count, or lymphopenia, a useful feature when treating autoimmune diseases but an undesirable side effect when activating $S1P_1$ in other cell types. Poirier *et al*. identified a G protein–biased $S1P_1$ agonist, SAR247799, which maintained endothelial function and reduced tissue damage caused by ischemia-reperfusion injury in two different animal models. The characteristics of SAR247799 show that it is possible to target $S1P_1$ in the endothelium without compromising immune responses. —WW

Sci. Signal. **12**, eaax8050 (2020).

Published by AAAS

REVIEW SUMMARY

COMPUTER SCIENCE

There's plenty of room at the Top: What will drive computer performance after Moore's law?

Charles E. Leiserson, Neil C. Thompson*, Joel S. Emer, Bradley C. Kuszmaul, Butler W. Lampson, Daniel Sanchez, Tao B. Schardl

BACKGROUND: Improvements in computing power can claim a large share of the credit for many of the things that we take for granted in our modern lives: cellphones that are more powerful than room-sized computers from 25 years ago, internet access for nearly half the world, and drug discoveries enabled by powerful supercomputers. Society has come to rely on computers whose performance increases exponentially over time.

Much of the improvement in computer performance comes from decades of miniaturization of computer components, a trend that was foreseen by the Nobel Prize–winning physicist Richard Feynman in his 1959 address, "There's Plenty of Room at the Bottom," to the American Physical Society. In 1975, Intel founder Gordon Moore predicted the regularity of this miniaturization trend, now called Moore's law, which, until recently, doubled the number of transistors on computer chips every 2 years.

Unfortunately, semiconductor miniaturization is running out of steam as a viable way to grow computer performance—there isn't much more room at the "Bottom." If growth in computing power stalls, practically all industries will face challenges to their productivity. Nevertheless, opportunities for growth in computing performance will still be available, especially at the "Top" of the computing-technology stack: software, algorithms, and hardware architecture.

ADVANCES: Software can be made more efficient by performance engineering: restructuring software to make it run faster. Performance engineering can remove inefficiencies in programs, known as software bloat, arising from traditional software-development strategies that aim to minimize an application's development time rather than the time it takes to run. Performance engineering can also tailor software to the hardware on which it runs, for example, to take advantage of parallel processors and vector units.

Algorithms offer more-efficient ways to solve problems. Indeed, since the late 1970s, the time to solve the maximum-flow problem improved nearly as much from algorithmic advances as from hardware speedups. But progress on a given algorithmic problem occurs unevenly and sporadically and must ultimately face diminishing returns. As such, we see the biggest benefits coming from algorithms for new problem domains (e.g., machine learning) and from developing new theoretical machine models that better reflect emerging hardware.

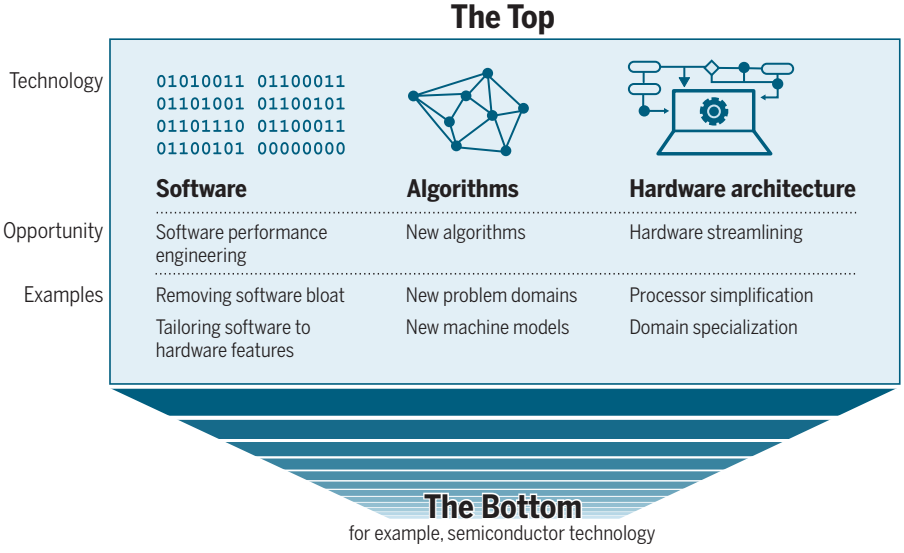
ON OUR WEBSITE

Read the full article at https://dx.doi.org/10.1126/science.aam9744

Hardware architectures can be streamlined—for instance, through processor simplification, where a complex processing core is replaced with a simpler core that requires fewer transistors. The freed-up transistor budget can then be redeployed in other ways—for example, by increasing the number of processor cores running in parallel, which can lead to large efficiency gains for problems that can exploit parallelism. Another form of streamlining is domain specialization, where hardware is customized for a particular application domain. This type of specialization jettisons processor functionality that is not needed for the domain. It can also allow more customization to the specific characteristics of the domain, for instance, by decreasing floating-point precision for machine-learning applications.

In the post-Moore era, performance improvements from software, algorithms, and hardware architecture will increasingly require concurrent changes across other levels of the stack. These changes will be easier to implement, from engineering-management and economic points of view, if they occur within big system components: reusable software with typically more than a million lines of code or hardware of comparable complexity. When a single organization or company controls a big component, modularity can be more easily reengineered to obtain performance gains. Moreover, costs and benefits can be pooled so that important but costly changes in one part of the big component can be justified by benefits elsewhere in the same component.

OUTLOOK: As miniaturization wanes, the silicon-fabrication improvements at the Bottom will no longer provide the predictable, broad-based gains in computer performance that society has enjoyed for more than 50 years. Software performance engineering, development of algorithms, and hardware streamlining at the Top can continue to make computer applications faster in the post-Moore era. Unlike the historical gains at the Bottom, however, gains at the Top will be opportunistic, uneven, and sporadic. Moreover, they will be subject to diminishing returns as specific computations become better explored. ■



Performance gains after Moore's law ends. In the post-Moore era, improvements in computing power will increasingly come from technologies at the "Top" of the computing stack, not from those at the "Bottom", reversing the historical trend.

CREDIT: N. CARY/*SCIENCE*

The list of author affiliations is available in the full article online.
*Corresponding author. Email: neil_t@mit.edu
Cite this article as C. E. Leiserson *et al.*, *Science* **368**, eaam9744 (2020). DOI: 10.1126/science.aam9744

REVIEW

COMPUTER SCIENCE

There's plenty of room at the Top: What will drive computer performance after Moore's law?

Charles E. Leiserson[1], Neil C. Thompson[1,2]*, Joel S. Emer[1,3], Bradley C. Kuszmaul[1]†, Butler W. Lampson[1,4], Daniel Sanchez[1], Tao B. Schardl[1]

The miniaturization of semiconductor transistors has driven the growth in computer performance for more than 50 years. As miniaturization approaches its limits, bringing an end to Moore's law, performance gains will need to come from software, algorithms, and hardware. We refer to these technologies as the "Top" of the computing stack to distinguish them from the traditional technologies at the "Bottom": semiconductor physics and silicon-fabrication technology. In the post-Moore era, the Top will provide substantial performance gains, but these gains will be opportunistic, uneven, and sporadic, and they will suffer from the law of diminishing returns. Big system components offer a promising context for tackling the challenges of working at the Top.

Over the past 50 years, the miniaturization of semiconductor devices has been at the heart of improvements in computer performance, as was foreseen by physicist Richard Feynman in his 1959 address (*1*) to the American Physical Society, "There's Plenty of Room at the Bottom." Intel founder Gordon Moore (*2*) observed a steady rate of miniaturization and predicted (*3*) that the number of transistors per computer chip would double every 2 years—a cadence, called Moore's law, that has held up considerably well until recently. Moreover, until about 2004, new transistors were not only smaller, they were also faster and more energy efficient (*4*), providing computers with ever more speed and storage capacity. Moore's law has driven economic progress ubiquitously.

Unfortunately, Feynman's "room at the bottom" is no longer plentiful. The *International Technology Roadmap for Semiconductors* [(*5*), p. 36] foresees an end to miniaturization, and Intel [(*6*), p. 14], a leader in microprocessor technology, has acknowledged an end to the Moore cadence. Indeed, Intel produced its 14-nm technology in 2014, but it stalled on producing its 10-nm technology, due in 2016, until 2019 (*7*). Although other manufacturers continued to miniaturize—for example, with the Samsung Exynos 9825 (*8*) and the Apple A13 Bionic (*9*)—they also failed to meet the Moore cadence. There isn't much more room at the bottom.

Why is miniaturization stalling? It's stalling because of fundamental physical limits—the physics of materials changes at atomic levels—and because of the economics of chip manufacturing. Although semiconductor technology may be able to produce transistors as small as 2 nm (20 Å), as a practical matter, miniaturization may end around 5 nm because of diminishing returns (*10*). And even if semiconductor technologists can push things a little further, the cost of doing so rises precipitously as we approach atomic scales (*11*, *12*).

In this review, we discuss alternative avenues for growth in computer performance after Moore's law ends. We believe that opportunities can be found in the higher levels of the computing-technology stack, which we refer to as the "Top." Correspondingly, by "Bottom" we mean the semiconductor technology that improved so dramatically during the Moore era. The layers of the computing stack harness the transistors and other semiconductor devices at the Bottom into useful computation at the Top to solve real-world problems. We divide the Top into three layers: (i) hardware architecture—programmable digital circuits that perform calculations; (ii) software—code that instructs the digital circuits what to compute; and (iii) algorithms—efficient problem-solving routines that organize a computation. We contend that even if device technologies at the Bottom cease to deliver performance gains, the Top will continue to offer opportunities.

Unlike Moore's law, which has driven up performance predictably by "lifting all boats," working at the Top to obtain performance will yield opportunistic, uneven, and sporadic gains, typically improving just one aspect of a particular computation at a time. For any given problem, the gains will suffer from the law of diminishing returns. In the long run, gains will depend on applying computing to new problems, as has been happening since the dawn of digital computers.

Working at the Top to obtain performance also differs from the Bottom in how it affects a computing system overall. The performance provided by miniaturization has not required substantial changes at the upper levels of the computing stack, because the logical behavior of the digital hardware, software, and data in a computation is almost entirely independent of the size of the transistors at the Bottom. As a result, the upper levels can take advantage of smaller and faster transistors with little or no change. By contrast—and unfortunately—many parts of the Top are dependent on each other, and thus when one part is restructured to improve performance, other parts must often adapt to exploit, or even tolerate, the changes. When these changes percolate through a system, it can take considerable human effort to correctly implement and test them, which increases both costs and risks. Historically, the strategies at the Top for improving performance coexisted with Moore's law and were used to accelerate particular applications that needed more than the automatic performance gains that Moore's law could provide.

Here, we argue that there is plenty of room at the Top, and we outline promising opportunities within each of the three domains of software, algorithms, and hardware. We explore the scale of improvements available in these areas through examples and data analyses. We also discuss why "big system components" will provide a fertile ground for capturing these gains at the Top.

Software

Software development in the Moore era has generally focused on minimizing the time it takes to develop an application, rather than the time it takes to run that application once it is deployed. This strategy has led to enormous inefficiencies in programs, often called software bloat. In addition, much existing software fails to take advantage of architectural features of chips, such as parallel processors and vector units. In the post-Moore era, software performance engineering—restructuring software to make it run faster—can help applications run more quickly by removing bloat and by tailoring software to specific features of the hardware architecture.

To illustrate the potential gains from performance engineering, consider the simple problem of multiplying two 4096-by-4096 matrices. Let us start with an implementation coded in Python, a popular high-level programming language. Here is the four-line kernel of the Python 2 code for matrix-multiplication:

```
for i in xrange(4096):
  for j in xrange(4096):
    for k in xrange(4096):
      C[i][j] += A[i][k] * B[k][j]
```

The code uses three nested loops and follows the method taught in basic linear-algebra

[1]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA. [2]MIT Initiative on the Digital Economy, Cambridge, MA, USA. [3]NVIDIA Research, Westford, MA, USA. [4]Microsoft Research, Cambridge, MA, USA.
*Corresponding author. Email: neil_t@mit.edu
†Present address: Google, Cambridge, MA, USA.

Table 1. Speedups from performance engineering a program that multiplies two 4096-by-4096 matrices. Each version represents a successive refinement of the original Python code. "Running time" is the running time of the version. "GFLOPS" is the billions of 64-bit floating-point operations per second that the version executes. "Absolute speedup" is time relative to Python, and "relative speedup," which we show with an additional digit of precision, is time relative to the preceding line. "Fraction of peak" is GFLOPS relative to the computer's peak 835 GFLOPS. See Methods for more details.

Version	Implementation	Running time (s)	GFLOPS	Absolute speedup	Relative speedup	Fraction of peak (%)
1	Python	25,552.48	0.005	1	—	0.00
2	Java	2,372.68	0.058	11	10.8	0.01
3	C	542.67	0.253	47	4.4	0.03
4	Parallel loops	69.80	1.969	366	7.8	0.24
5	Parallel divide and conquer	3.80	36.180	6,727	18.4	4.33
6	plus vectorization	1.10	124.914	23,224	3.5	14.96
7	plus AVX intrinsics	0.41	337.812	62,806	2.7	40.45

classes. It turns out, however, that this naïve code leaves much of the performance available on modern computers "on the table." The code takes about 7 hours on a modern computer to compute the matrix product, as shown by the first row (version 1) in Table 1, achieving only 0.0006% of the peak performance of the machine. (Incidentally, Python 3 requires about 9 hours for the same computation.)

How can this naïve matrix-multiplication code be performance engineered? Simply choosing a more efficient programming language speeds up this calculation dramatically. For example, coding it in Java (version 2) produces a speedup of 10.8×, and coding it in C (version 3) produces an additional speedup of 4.4×, yielding an execution time that is 47 times faster than the original Python. This performance improvement comes from reducing the number of operations the program performs. In particular, Java and C avoid the extraneous work that Python does under the hood to make programming easier. The price for this performance gain is programmer productivity: Coding in C is more onerous than coding in Python, and Java lies somewhere in between.

Although switching languages gains a speedup of almost 50×, tailoring the matrix code to exploit specific features of the hardware makes it run an additional 1300 times faster. This gain comes from parallelizing the code to run on all 18 of the processing cores (version 4), exploiting the processor's memory hierarchy (version 5), vectorizing the code (version 6), and using Intel's special Advanced Vector Extensions (AVX) instructions (version 7). The final optimized code performs the task in only 0.41 s—more than 60,000 times faster than the 7 hours of the original Python code!

The point of this example is to illustrate the potential gains available from performance engineering naïvely coded software. In the particular case of matrix multiplication, a good programmer could avoid this programming effort by using optimized code from existing software libraries. If she were writing code to solve a new problem, however, she would need to optimize the code herself. And although not every application can improve by nearly five orders of magnitude through performance engineering, most modern software systems contain ample opportunity for performance enhancement, especially if the codebase is large enough.

During the post-Moore era, it will become ever more important to make code run fast and, in particular, to tailor it to the hardware on which it runs. Modern computers provide architectural features designed to make code run fast. For example, versions 4 and 6 exploit parallelism, which is the ability of computers to perform multiple operations at the same time. Version 5 exploits locality, which is the computer's ability to access data elements efficiently when they are collocated in memory (spatial locality) or have been accessed recently (temporal locality). Version 7 exploits both parallelism and locality through carefully coordinated use of Intel's AVX instructions. As we shall see in the Hardware architecture section, architectures are likely to become increasingly heterogeneous, incorporating both general-purpose and special-purpose circuitry. To improve performance, programs will need to expose more parallelism and locality for the hardware to exploit. In addition, software performance engineers will need to collaborate with hardware architects so that new processors present simple and compelling abstractions that make it as easy as possible to exploit the hardware.

Beyond the tailoring of software to hardware is the question of bloat: Where does software bloat come from? Certainly, some bloat comes from trading off efficiency for other desirable traits, such as coding ease, as versions 1 to 3 of the matrix-multiplication code illustrate. Bloat also comes from a failure to tailor code to the underlying architecture, as versions 4 to 7 show. But much software bloat arises from software-development strategies (*13*, *14*), such as reduction.

The idea of reduction is this. Imagine that you are a programmer who has been given a problem A to solve (for example, distinguishing between a yes or no spoken response). You could write specialized code to solve A directly, but instead, you might notice that a related problem B has already been solved (existing speech-recognition software that understands many words, including yes and no). It will take you far less effort to solve A by converting it into a problem that can be solved with the existing code for B, that is, by reducing A to B.

Inefficiencies can arise both from the reduction itself (translating A to B) and from the generality of B (the solution to B is not tailored specifically to A). But the largest bloat arises from the compounding of reductions: reducing A to B, B to C, C to D, and so on. Even if each reduction achieves an impressive 80% efficiency, a sequence of two independent reductions achieves just 80% × 80% = 64%. Compounding 20 more times yields an efficiency of less than 1%, or 100× in bloat.

Because of the accumulated bloat created by years of reductionist design during the Moore era, there are great opportunities to make programs run faster. Unfortunately, directly solving problem A using specialized software requires expertise both in the domain of A and in performance engineering, which makes the process more costly and risky than simply using reductions. The resulting specialized software to solve A is often more complex than the software that reduces A to B. For example, the fully optimized code in Table 1 (version 7) is more than 20 times longer than the source code for the original Python version (version 1).

Indeed, simple code tends to be slow, and fast code tends to be complicated. To create a world where it is easy to write fast code, application programmers must be equipped with the knowledge and skills to performance-engineer

their code, and productivity tools to assist must be improved considerably.

Abstractly, software performance engineering can be viewed as a simple process involving a single loop: (i) Measure the performance of program A. (ii) Make a change to program A to produce a hopefully faster program A'. (iii) Measure the performance of program A'. (iv) If A' beats A, set $A = A'$. (v) If A is still not fast enough, go to (ii). But today, our systems are sufficiently complicated that measurements must often be repeated many times to develop confidence that one version of a program outperforms another.

As hardware becomes increasingly specialized and heterogeneous (see the Hardware architecture section), high-performing code will become even more difficult to write. Because faster software will be increasingly important for better performance in the post-Moore era, however, the computing industry, researchers, and government should all be well motivated to develop performance-engineering technologies.

Algorithms

Algorithmic advances have already made many contributions to performance growth and will continue to do so in the future. A major goal is to solve a problem with less computational work. For example, by using Strassen's algorithm (*15*) for matrix multiplication, the highly optimized code in version 7 of Table 1 can be improved by about 10%. For some problems, the gains can be much more impressive: The President's Council of Advisors on Science and Technology concluded in 2010 that "*performance gains due to improvements in algorithms have vastly exceeded even the dramatic performance gains due to increased processor speed*" (emphasis theirs) [(*16*), p. 71].

Because algorithm design requires human ingenuity, however, it is hard to anticipate advances. To illustrate the nature of algorithmic progress, consider the classical operations-research problem of finding the maximum flow in a network [(*17*), chap. 26], which can be used to model the movement of traffic in a road network, blood through the circulatory system, or electricity in a circuit. Linear programming is a straightforward way to solve the maximum-flow problem, but the 20 years between 1975 and 1995 saw a sequence of algorithmic innovations that greatly improved on it.

Figure 1 shows the progress in maximum-flow algorithms over time. The performance gain of the best algorithm has rivaled the gain due to Moore's law over the 38 years of the data (just over four orders of magnitude versus just under five), even though no new algorithm has improved the performance of this particular problem over the past 20 years. This example highlights three salient observations about algorithms: (i) Progress on a given algorithmic problem occurs unevenly and sporadically. (ii) The benefits from algorithmic innovation can rival gains from Moore's law. (iii) The algorithmic improvements in solving any given problem must eventually diminish.

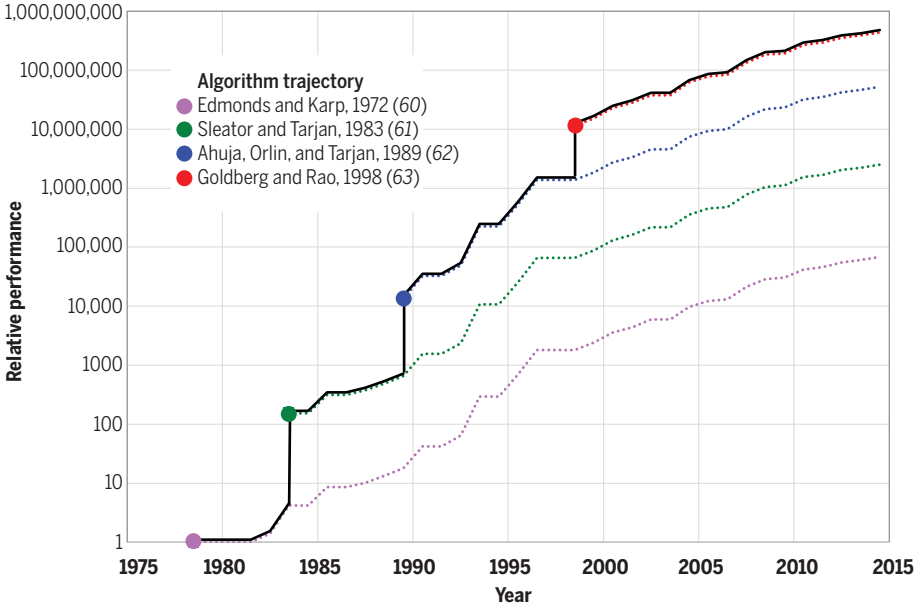


Fig. 1. Major algorithmic advances in solving the maximum-flow problem on a graph with $n = 10^{12}$ vertices and $m = n^{1.1}$ edges. The vertical axis shows how many problems (normalized to the year 1978) could theoretically be solved in a fixed time on the best microprocessor system available in that year. Each major algorithm is shown as a circle in the year of its invention, except the first, which was the best algorithm in 1978. The dotted trajectory shows how faster computers [as measured by SPECint scores in Stanford's CPU database (*56*)] make each algorithm faster over time. The solid black line shows the best algorithm using the best computer at any point in time. Each algorithm's performance is measured by its asymptotic complexity, as described in the Methods.

Because this example focuses on a well-known problem, however, it misses a key aspect of how algorithmic performance engineering can speed up computing: by providing efficient solutions to new problems. For example, more than a quarter of the 146 papers at the 2016 Association for Computing Machinery (ACM) Symposium on Discrete Algorithms focus on problems that had not previously been studied algorithmically. Consequently, although research on old problems may still yield marginal gains, much of the progress in algorithms will come from three sources: (i) attacking new problem domains, (ii) addressing scalability concerns, and (iii) tailoring algorithms to take advantage of modern hardware. We discuss each source in turn.

New problem domains continually create a need for new algorithms. Domains such as machine learning, social networks, security, robotics, game theory, sensor networks, and video coding were tiny or nonexistent 30 years ago but are now economically important enough to demand efficient algorithms. Many companies have gained a competitive advantage thanks to algorithms. Google's PageRank algorithm (*18*) made its World Wide Web search superior, and the auction algorithms of Google AdWords (*19*), which allow advertisers to bid for display space based on users' search terms, made it highly profitable. Content-delivery networks, which delivered more than half of internet traffic in 2016 (*20*), depend on efficient algorithms to avoid congestion. Many sciences also depend on good algorithms. For example, DNA sequencing in computational biology depends on efficient dynamic-programming algorithms (*21*).

Moore's law has enabled today's high-end computers to store over a terabyte of data in main memory, and because problem sizes have grown correspondingly, efficient algorithms are needed to make solutions affordable. Sublinear algorithms (*22*, *23*) provide one example of how to address problems of scale. For instance, to find the median of a trillion numbers, it takes at least a trillion memory operations just to read the input data. But many problems do not need the exact median and work perfectly well with only a good estimate of the median. For these problems, we can instead extract a random sample of, say, a million numbers and compute the median of that sample. The result is a highly accurate estimate of the true median that can be computed a million times faster. The field of algorithms is full of strategies for dealing with scalability.

Tailoring an algorithm to exploit modern hardware can make it much faster (Table 1). Nevertheless, most algorithms today are still

designed using the serial random-access machine model (*24*) originally developed in the 1960s and 1970s, which assumes that a processor can do only one operation at a time and that the cost to access any part of the memory is the same. Such algorithms often use modern hardware inefficiently because they underutilize the machine's many parallel-processing cores and vector units, each of which can perform many operations per clock cycle, and they fail to exploit caching, which can speed up data accesses by two orders of magnitude.

Although algorithms research has developed mathematical models for salient features of modern computers, such as parallel and vector processing (*25*–*32*) and cache hierarchies (*33*–*35*), a substantial gap between algorithm and implementation remains. Part of the problem is that each model tends to address just one aspect—such as parallelism, vector units, or caching—and yet tailoring an algorithm to a modern computer requires an understanding of all of them. Moreover, in an effort to gain every bit of performance, some hardware features—such as simultaneous multithreading, dynamic voltage and frequency scaling, direct-mapped caches, and various special-purpose instructions—actually make it more difficult to tailor algorithms to hardware, because they cause variability and unpredictability that simple theoretical models cannot easily capture.

One possible solution is autotuning (*36*, *37*), which searches a parametrized space of possible implementations to find the fastest one. With modern machine learning, it may even be possible to include implementations that differ by more than the values of a few parameters. Unfortunately, autotuning and machine learning tend to be too time consuming to ask that every algorithm incur this large up-front cost. Furthermore, these approaches actually make algorithm design harder, because the designer cannot easily understand the ramifications of a design choice. In the post-Moore era, it will be essential for algorithm designers and hardware architects to work together to find simple abstractions that designers can understand and that architects can implement efficiently.

Hardware architecture

Historically, computer architects used more and more transistors to make serial computations run faster, vastly increasing the complexity of processing cores, even though gains in performance suffered from diminishing returns over time (*38*). We argue that in the post-Moore era, architects will need to adopt the opposite strategy and focus on hardware streamlining: implementing hardware functions using fewer transistors and less silicon area.

As we shall see, the primary advantage of hardware streamlining comes from providing additional chip area for more circuitry to operate in parallel. Thus, the greatest benefit accrues to applications that have ample parallelism. Indeed, the performance of hardware for applications without much parallelism has already stagnated. But there is plenty of parallelism in many emerging application domains, such as machine learning, graphics, video and image processing, sensory computing, and signal processing. Computer architects should be able to design streamlined architectures to provide increasing performance for these and other domains for many years after Moore's law ends.

We can use historical data to observe the trend of architectural reliance on parallelism. Figure 2 plots three sets of benchmark data for microprocessors: SPECint performance (black squares and gray diamonds), SPECint-rate performance (black, orange, blue, and red squares), and microprocessor clock frequency (green dots). As the green dots in the figure show, clock speed increased by a factor of more than 200 from 1985 to 2005, when it plateaued owing to the end of Dennard scaling, which we shall discuss shortly. Driven by increasing clock speed and other architectural changes during the Dennard-scaling era, microprocessor performance rapidly improved, as measured by the SPECint and SPECint-rate benchmarks (black squares), which aim to model computer performance on typical user workloads (*39*). The SPECint benchmark consists of mostly serial code, whereas the SPECint-rate benchmark is parallel. The two benchmarks perform the same on single-processor computers. But after 2004, as machines added multiple cores and other forms of explicit parallelism, the two diverge. Indeed, the performance of parallel applications on the best-performing chips in each year (colored squares) grew by a factor of 30 from 2004 to 2015, improving on average by about a factor of two every 2 years. By contrast, over the same time period, the largely serial SPECint benchmark (gray diamonds) scaled up by only a factor of three.

Besides parallelism, an application needs locality to benefit from streamlining. As an example, when data are transferred from external dynamic random access memory (DRAM) memory chips to a processing chip, it should be used multiple times before being transferred back. For an application with little locality, increasing parallelism causes traffic to off-chip memory to increase proportionally and

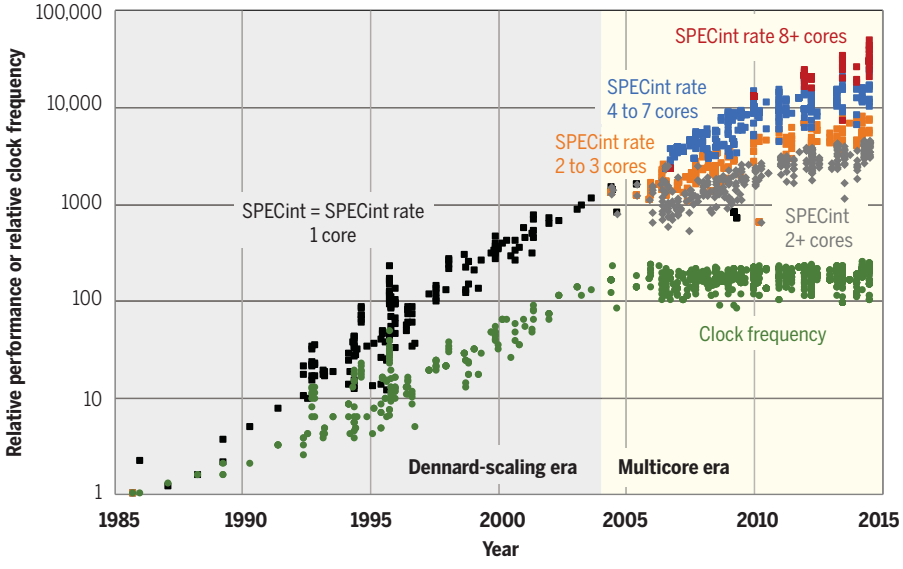


Fig. 2. SPECint (largely serial) performance, SPECint-rate (parallel) performance, and clock-frequency scaling for microprocessors from 1985 to 2015, normalized to the Intel 80386 DX microprocessor in 1985. Microprocessors and their clock frequencies were obtained from the Stanford CPU database (*56*). Microprocessor performance is measured in terms of scaled performance scores on the SPECint and SPECint-rate performance benchmarks obtained from (*39*). (See Methods for details.) Black squares identify single-core processors, for which SPECint and SPECint-rate benchmark performances are the same. Orange, blue, and red squares plot the SPECint-rate benchmark performance of various multicore processors, where orange squares identify processors with two to three cores, blue squares identify processors with four to seven cores, and red squares identify processors with eight or more cores. The gray diamonds plot the SPECint benchmark performance on multicore processors. The round green dots plot processor clock frequencies (also normalized to the Intel 80386). The gray background highlights the Dennard-scaling era (nominally up to 2004), and the white background highlights the multicore era (beyond 2004).

eventually exceeds the bandwidth of its channel to memory, so that the application is memory bound. For an application with good locality, however, as parallelism increases, the amount of off-chip memory traffic increases much more slowly, enabling all the chip's computing engines to do useful work without idling. Fortunately, many important application domains contain plenty of both locality and parallelism.

Hardware streamlining can exploit locality in other ways, especially for domain-specific processors, which we shall discuss shortly. For example, explicit data orchestration (*40*) exploits locality to increase the efficiency with which data are moved throughout the memory hierarchy [(*41*), chap. 4]. On-chip interconnects can become simpler and consume less power and area if the application using them contains locality. For example, systolic arrays (*42*) can perform matrix computations more efficiently using an area-efficient mesh interconnect than a general-purpose interconnect.

Although hardware will increase in capability because of streamlining, we do not think that average clock speed will increase after Moore's law ends, and it may in fact diminish slightly. Figure 2 shows that clock speed plateaued in 2005, when microprocessor design became power constrained. Before 2004, computer architects found ways to increase clock frequency without hitting hard power limits. Dennard scaling—reducing voltage as clock frequency increased—allowed processors to run faster without increasing power usage. (In practice, processor manufacturers often increased clock frequency without reducing voltage proportionally, which did increase chip power.) Since 2004, however, Moore's law has provided many more transistors per chip, but because the ability to power them has not grown appreciably (*43*), architects have been forced to innovate just to prevent clock rates from falling. Slightly lower clock frequency and supply voltage reduce the power per transistor enough that substantially more circuitry can run in parallel. If the workload has enough parallelism, the added computing more than compensates for the slower clock. Serial applications may see somewhat worse performance, but cleverness can reduce this cost. For example, Intel's "Turbo" mode [(*41*), p. 28], runs the clock faster when fewer cores are active. (Other techniques to reduce transistor switching include using more transistors in caches, power-gating unused circuitry, and minimizing signal switching.)

Now that designers have embraced parallelism, the main question will be how to streamline processors to exploit application parallelism. We expect two strategies to dominate: processor simplification and domain specialization.

Processor simplification (*44*) replaces a complex processing core with a simpler core that requires fewer transistors. A modern core contains many expensive mechanisms to make serial instruction streams run faster, such as speculative execution [(*41*), section 3.6], where the hardware guesses and pursues future paths of code execution, aborting and reexecuting if the guess is wrong. If a core can be simplified to occupy, say, half as many transistors, then twice as many cores can fit on the chip. For this trade-off to be worthwhile, the workload must have enough parallelism that the additional cores are kept busy, and the two simplified cores must do more useful computing than the single complex one.

Domain specialization (*11*, *43*, *45*) may be even more important than simplification. Hardware that is customized for an application domain can be much more streamlined and use many fewer transistors, enabling applications to run tens to hundreds of times faster (*46*). Perhaps the best example today is the graphics-processing unit (GPU) [(*41*), section 4.4], which contains many parallel "lanes" with streamlined processors specialized to computer graphics. GPUs deliver much more performance on graphics computations, even though their clocks are slower, because they can exploit much more parallelism. GPU logic integrated into laptop microprocessors grew from 15 to 25% of chip area in 2010 to more than 40% by 2017 (see Methods), which shows the importance of GPU accelerators. Moreover, according to the Top 500 website, which tracks high-performance computing technology, only about 10% of supercomputers that were added to the Top 500 list in the year 2012 contained accelerators (often GPUs), but by 2017, that share had grown to 38% (*12*). Hennessy and Patterson (*45*) foresee a move from general-purpose toward domain-specific architectures that run small compute-intensive kernels of larger systems for tasks such as object recognition or speech understanding. The key requirement is that the most expensive computations in the application domain have plenty of parallelism and locality.

A specialized processor is often first implemented as an attached device to a general-purpose processor. But the forces that encourage specialization must be balanced with the forces that demand broadening: expanding the functionality of the specialized processor to make it more autonomous from the general-purpose processor and more widely useful to other application domains (*47*).

The evolution of GPUs demonstrates this trade-off. GPUs were originally developed specifically for rendering graphics, and as a result, GPUs are next to useless for many other computational tasks, such as compiling computer programs or running an operating system. But GPUs have nevertheless broadened to be handy for a variety of nongraphical tasks, such as linear algebra. Consider the matrix-multiplication problem from the Software section. An Advanced Micro Devices (AMD) FirePro S9150 GPU (*48*) can produce the result in only 70 ms, which is 5.4 times faster than the optimized code (version 7) and a whopping 360,000 times faster than the original Python code (version 1).

As another example of the trade-off between broadening and specialization, GPUs were crucial to the "deep-learning" revolution (*49*), because they were capable of training large neural networks that general-purpose processors could not train (*50*, *51*) fast enough. But specialization has also succeeded. Google has developed a tensor-processing unit (TPU) (*52*) specifically designed for deep learning, embracing special-purpose processing and eschewing the broader functionality of GPUs.

During the Moore era, specialization usually yielded to broadening, because the return on investment for developing a special-purpose device had to be amortized by sales over the limited time before Moore's law produced a general-purpose processor that performs just as well. In the post-Moore era, however, we expect to see more special-purpose devices, because they will not have comparably performing general-purpose processors right around the corner to compete with. We also expect a diversity of hardware accelerators specialized for different application domains, as well as hybrid specialization, where a single device is tailored for more than one domain, such as both image processing and machine learning for self-driving vehicles (*53*). Cloud computing will encourage this diversity by aggregating demand across users (*12*).

Big components

In the post-Moore era, performance engineering, development of algorithms, and hardware streamlining will be most effective within big system components (*54*). A big component is reusable software with typically more than a million lines of code, hardware of comparable complexity, or a similarly large software-hardware hybrid. This section discusses the technical and economic reasons why big components are a fertile ground for obtaining performance at the Top.

Changes to a system can proceed without much coordination among engineers, as long as the changes do not interfere with one another. Breaking code into modules and hiding its implementation behind an interface make development faster and software more robust (*55*). Modularity aids performance engineering, because it means that code within a module can be improved without requiring the rest of the system to adapt. Likewise, modularity aids in hardware streamlining, because the hardware can be restructured

without affecting application programming interfaces (APIs). Performance engineering and hardware streamlining that do not require coordination already occur this way, and we expect them to continue to do so.

Many of the most valuable future opportunities for improving performance will not arise locally within a single module, however, but from broad-based and systematic changes across many modules affecting major parts of a system. For example, because so much software is reductionist, the larger a system, the greater the opportunity for huge performance gains when layer upon layer of reductions are replaced by a lean and direct implementation. But large-scale reengineering that affects many modules requires the coordination of many engineers, which is costly and risky.

Big system components (web browsers, database systems, operating systems, microprocessors, GPUs, compilers, disk-storage systems, and so on) offer both technical opportunities for making applications run fast and a context where it is economically viable to take advantage of them. As an example of the types of changes that can be made within a big component, consider a microprocessor. An instruction-set architecture (ISA) is the interface by which software tells the processor what to do. Manufacturers routinely make major internal changes to improve performance without changing the ISA so that old software continues to run correctly. For example, the Intel 8086 released in 1978 had 29,000 transistors (*56*), whereas the 22-core Xeon E5-2699 v4, released in 2016, has about 248,000 times more transistors (*57*) that produce more than a million times better performance, according to SPECint rate (*39*). In that time, the ISA has grown by less than a factor of four (*58*), and old programs continue to work even as the interface adds new functionality.

Big components provide good opportunities for obtaining performance, but opportunity alone is not enough. To outweigh the costs and risks of making changes, an economic incentive must exist. If a commercial enterprise or a nonprofit organization owns a big component, it can justify the investment to enhance performance because it reaps the benefit when the job is done. Single ownership also helps with coordination costs. If many parties must agree to make a change, but it takes only a few to veto it, then it can be hard for change to happen. Even the fear of high coordination costs can block change in a large codebase if too many parties are involved. But when a single entity owns a big component, it has the power to make vast changes and pool the costs and benefits. It can choose to reengineer as many modules as it can justify economically and coordinate with the outside world only at the big component's interface.

Conclusions

As miniaturization wanes, the silicon-fabrication improvements at the Bottom will no longer provide the predictable, broad-based gains in computer performance that society has enjoyed for more than 50 years. Performance-engineering of software, development of algorithms, and hardware streamlining at the Top can continue to make computer applications faster in the post-Moore era, rivaling the gains accrued over many years by Moore's law. Unlike the historical gains at the Bottom, however, the gains at the Top will be opportunistic, uneven, sporadic, and subject to diminishing returns as problems become better explored. But even where opportunities exist, it may be hard to exploit them if the necessary modifications to a component require compatibility with other components. Big components can allow their owners to capture the economic advantages from performance gains at the Top while minimizing external disruptions.

In addition to the potential at the Top, nascent technologies—such as 3D stacking, quantum computing, photonics, superconducting circuits, neuromorphic computing, and graphene chips—might provide a boost from the Bottom. At the moment, these technologies are in their infancy and lack the maturity to compete with today's silicon-based semiconductor technology in the near future (*59*). Although we applaud investment in these technologies at the Bottom because of their long-term potential, we view it as far more likely that, at least in the near term, performance gains for most applications will originate at the Top.

Methods

Table 1

Each running time is the minimum of five runs on an Amazon AWS c4.8xlarge spot instance, a dual-socket Intel Xeon E5-2666 v3 system with a total of 60 gibibytes of memory. Each Xeon is a 2.9-GHz 18-core CPU with a shared 25-mebibyte L3-cache. Each processor core has a 32–kibibyte (KiB) private L1-data-cache and a 256-KiB private L2-cache. The machine was running Fedora 22, using version 4.0.4 of the Linux kernel. The Python version was executed using Python 2.7.9. The Java version was compiled and run using OpenJDK version 1.8.0_51. All other versions were compiled using GNU Compiler Collection (GCC) 5.2.1 20150826.

Figure 1

Each curve models how hardware improvements grow the performance of a fixed maximum-flow algorithm, starting from the year the algorithm was published. The performance of each algorithm was measured as the number of graphs of $n = 10^{12}$ vertices, $m = n^{1.1} \sim 15.8 \times 10^{12}$ edges, and 64-bit integer capacities that can be solved in a fixed amount of time, normalized such that the first point starts at 1. We calculated the performance of each algorithm based on its asymptotic complexity in terms of big-O notation, assuming that the constant hidden by the big-O is 1. We excluded approximate algorithms from consideration, because these algorithms do not necessarily return the same qualitative answer. We also excluded algorithms whose stated asymptotic bounds ignore logarithmic factors to simplify the performance comparisons. We identified four maximum-flow algorithms developed since 1978 that each delivered a performance improvement of more than a factor of four over its predecessor: Edmonds and Karp's $O(m^2 \log U)$ algorithm (*60*); Sleator and Tarjan's $O(mn \log n)$ algorithm (*61*); Ahuja, Orlin, and Tarjan's $O(mn \log(n\sqrt{\log U}/m + 2))$ algorithm (*62*); and Goldberg and Rao's $O(n^{2/3} m \log(n^2/m) \log U)$ algorithm (*63*). Each curve plots the best performance achieved by any processor up to that date, on the basis of scaled MIPS (million instructions per second) estimates and SPECint scores recorded in Stanford's CPU database (*56*). These processor performance measurements are normalized using the same method as in Fig. 2.

Figure 1 ignores the constant factors in the performance of maximum-flow algorithms because comparably engineered implementations of these algorithms are not available. Nevertheless, the effects from changes in constant factors between algorithms can be inferred from the plot. For example, if a constant factor were 10 times larger between one algorithm and a later one, then the curve for the later algorithm would be one order of magnitude lower (i.e., one division less on the y axis).

Figure 2

Performance measurements for different processors were gathered from Stanford's CPU database (*56*). For a processor released before 1992, its performance is measured as the recorded estimate of how many MIPS that processor can perform. For subsequent processors, each performance measurement corresponds to the best recorded score for that processor on either the SPECint1992, SPECint1995, SPECint2000, or SPECint2006 benchmarks. Performance for multicore processors is measured using both standard SPECint scores (gray diamonds) and in terms of throughput (multicolored squares), using SPECint2006 rate scores. All SPECint performance scores were obtained from (*39*). The date for an Intel processor released after 2004 is plotted as the first day in the quarter when that processor was released, unless a more precise release date was found. Scores for different SPECint benchmarks were normalized by scaling factors, which were computed from the geometric-mean ratios of scores for consecutive versions of SPECint for processors that were evaluated on both versions. The MIPS estimates were normalized with

the SPECint scores such that a score of 1 on SPECint1992 corresponds to 1 MIPS.

GPU logic integrated into laptop microprocessors

We obtained, from WikiChip (*57*), annotated die photos for Intel microprocessors with GPUs integrated on die, which began in 2010 with Sandy Bridge. We measured the area in each annotated photo dedicated to a GPU and calculated the ratio of this area to the total area of the chip. Intel's quad-core chips had approximately the following percentage devoted to the GPU: Sandy Bridge (18%), Ivy Bridge (33%), Haswell (32%), Skylake (40 to 60%, depending on version), Kaby Lake (37%), and Coffee Lake (36%). Annotated die photos for Intel microarchitectures newer than Coffee Lake were not available and therefore not included in the study. We did not find enough information about modern AMD processors to include them in this study.

REFERENCES AND NOTES

1. R. P. Feynman, There's plenty of room at the bottom. *Eng. Sci.* **23**, 22–36 (1960).
2. G. E. Moore, Cramming more components onto integrated circuits. *Electronics* **38**, 1–4 (1965).
3. G. E. Moore, "Progress in digital integrated electronics" in *International Electron Devices Meeting Technical Digest* (IEEE, 1975), pp. 11–13.
4. R. H. Dennard *et al*., Design of ion-implanted MOSFET's with very small physical dimensions. *JSSC* **9**, 256–268 (1974).
5. ITRS, International Technology Roadmap for Semiconductors 2.0, executive report (2015); www.semiconductors.org/wp-content/uploads/2018/06/0_2015-ITRS-2.0-Executive-Report-1.pdf.
6. Intel Corporation, Form 10K (annual report). SEC filing (2016); www.sec.gov/Archives/edgar/data/50863/000005086316000105/a10kdocument12262015q4.htm.
7. I. Cutress, Intel's 10nm Cannon Lake and Core i3-8121U deep dive review (2019); www.anandtech.com/show/13405/intel-10nm-cannon-lake-and-core-i3-8121u-deep-dive-review.
8. K. Hinum, Samsung Exynos 9825 (2019); www.notebookcheck.net/Samsung-Exynos-9825-SoC-Benchmarks-and-Specs.432496.0.html.
9. K. Hinum, Apple A13 Bionic (2019); www.notebookcheck.net/Apple-A13-Bionic-SoC.434834.0.html.
10. R. Merritt, "Path to 2 nm may not be worth it," *EE Times*, 23 March 2018; www.eetimes.com/document.asp?doc_id=1333109.
11. R. Colwell, "The chip design game at the end of Moore's Law," presented at Hot Chips, Palo Alto, CA, 25 to 27 August 2013.
12. N. C. Thompson, S. Spanuth, The decline of computers as a general-purpose technology: Why deep learning and the end of Moore's Law are fragmenting computing. SSRN 3287769 [Preprint]. 20 November 2019; doi: 10.2139/ssrn.3287769
13. J. Larus, Spending Moore's dividend. *Commun. ACM* **52**, 62–69 (2009). doi: 10.1145/1506409.1506425
14. G. Xu, N. Mitchell, M. Arnold, A. Rountev, G. Sevitsky, "Software bloat analysis: Finding, removing, and preventing performance problems in modern large-scale object-oriented applications" in *FoSER '10: Proceedings of the FSE/SDP Workshop on Future of Software Engineering Research* (ACM, 2010), pp. 421–426.
15. V. Strassen, Gaussian elimination is not optimal. *Numer. Math.* **13**, 354–356 (1969). doi: 10.1007/BF02165411
16. President's Council of Advisors on Science and Technology, "Designing a digital future: Federally funded research and development in networking and information technology" (Technical report, Executive Office of the President, 2010); www.cis.upenn.edu/~mkearns/papers/nitrd.pdf.
17. T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, *Introduction to Algorithms* (MIT Press, ed. 3, 2009).
18. S. Brin, L. Page, The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.* **30**, 107–117 (1998). doi: 10.1016/S0169-7552(98)00110-X
19. A. Mehta, A. Saberi, U. Vazirani, V. Vazirani, Adwords and generalized online matching. *J. Assoc. Comput. Mach.* **54**, 22 (2007). doi: 10.1145/1284320.1284321
20. Cisco Systems, Inc., "Cisco visual networking index (VNI): Complete forecast update, 2017–2022," Presentation 1465272001663118, Cisco Systems, Inc., San Jose, CA, December 2018; https://web.archive.org/web/20190916132155/https://www.cisco.com/c/dam/m/en_us/network-intelligence/service-provider/digital-transformation/knowledge-network-webinars/pdfs/1211_BUSINESS_SERVICES_CKN_PDF.pdf.
21. D. Gusfield, *Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology* (Cambridge Univ. Press, 1997).
22. R. Kumar, R. Rubinfeld, Algorithms column: Sublinear time algorithms. *SIGACT News* **34**, 57–67 (2003). doi: 10.1145/954092.954103
23. R. Rubinfeld, A. Shapira, Sublinear time algorithms. *SIDMA* **25**, 1562–1588 (2011). doi: 10.1137/100791075
24. A. V. Aho, J. E. Hopcroft, J. D. Ullman, *The Design and Analysis of Computer Algorithms* (Addison-Wesley Publishing Company, 1974).
25. R. L. Graham, Bounds for certain multiprocessing anomalies. *Bell Syst. Tech. J.* **45**, 1563–1581 (1966). doi: 10.1002/j.1538-7305.1966.tb01709.x
26. R. P. Brent, The parallel evaluation of general arithmetic expressions. *J. Assoc. Comput. Mach.* **21**, 201–206 (1974). doi: 10.1145/321812.321815
27. S. Fortune, J. Wyllie, "Parallelism in random access machines" in *STOC '78: Proceedings of the 10th Annual ACM Symposium on Theory of Computing* (ACM, 1978), pp. 114–118.
28. R. M. Karp, V. Ramachandran, "Parallel algorithms for shared-memory machines" in *Handbook of Theoretical Computer Science: Volume A, Algorithms and Complexity* (MIT Press, 1990), chap. 17, pp. 869–941.
29. G. E. Blelloch, *Vector Models for Data-Parallel Computing* (MIT Press, 1990).
30. L. G. Valiant, A bridging model for parallel computation. *Commun. ACM* **33**, 103–111 (1990). doi: 10.1145/79173.79181
31. D. Culler *et al*., "LogP: Towards a realistic model of parallel computation" in *Fourth ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming* (ACM, 1993), pp. 1–12.
32. R. D. Blumofe, C. E. Leiserson, Space-efficient scheduling of multithreaded computations. *SIAM J. Comput.* **27**, 202–229 (1998). doi: 10.1137/S0097539793259471
33. J. S. Vitter, Algorithms and data structures for external memory. *Found. Trends Theor. Comput. Sci.* **2**, 305–474 (2008). doi: 10.1561/0400000014
34. J.-W. Hong, H. T. Kung, "I/O complexity: The red-blue pebble game" in *STOC '81: Proceedings of the 13th Annual ACM Symposium on Theory of Computing* (ACM, 1981), pp. 326–333.
35. M. Frigo, C. E. Leiserson, H. Prokop, S. Ramachandran, "Cache-oblivious algorithms" in *FOCS '99: Proceedings of the 40th Annual Symposium on Foundations of Computer Science* (IEEE, 1999), pp. 285–297.
36. M. Frigo, A fast Fourier transform compiler. *ACM SIGPLAN Not.* **34**, 169–180 (1999). doi: 10.1145/301631.301661
37. J. Ansel *et al*., "OpenTuner: An extensible framework for program autotuning" in *PACT '14: 2014 23rd International Conference on Parallel Architecture and Compilation Techniques* (ACM, 2014), pp. 303–316.
38. S. Borkar, "Thousand core chips: A technology perspective" in *DAC '07: Proceedings of the 44th Annual Design Automation Conference* (ACM, 2007), pp. 746–749.
39. Standard Performance Evaluation Corporation, SPEC CPU 2006 (2017); www.spec.org/cpu2006.
40. M. Pellauer *et al*., "Buffets: An efficient and composable storage idiom for explicit decoupled data orchestration" in *ASPLOS '19: Proceedings of the 24th International Conference on Architectural Support for Programming Languages and Operating Systems* (ACM, 2019), pp. 137–151.
41. J. L. Hennessy, D. A. Patterson, *Computer Architecture: A Quantitative Approach* (Morgan Kaufmann, ed. 6, 2019).
42. H. T. Kung, C. E. Leiserson, "Systolic arrays (for VLSI)" in *Sparse Matrix Proceedings 1978*, I. S. Duff, G. W. Stewart, Eds. (SIAM, 1979), pp. 256–282.
43. M. B. Taylor, "Is dark silicon useful?: Harnessing the four horsemen of the coming dark silicon apocalypse" in *DAC '12: Proceedings of the 49th Annual Design Automation Conference* (ACM, 2012), pp. 1131–1136.
44. A. Agarwal, M. Levy, "The kill rule for multicore" in *DAC '07: Proceedings of the 44th Annual Design Automation Conference* (ACM, 2007), pp. 750–753.
45. J. L. Hennessy, D. A. Patterson, A new golden age for computer architecture. *Commun. ACM* **62**, 48–60 (2019). doi: 10.1145/3282307
46. R. Hameed *et al*., "Understanding sources of inefficiency in general-purpose chips" in *ISCA '10: Proceedings of the 37th Annual International Symposium on Computer Architecture* (ACM, 2010), pp. 37–47.
47. T. H. Myer, I. E. Sutherland, On the design of display processors. *Commun. ACM* **11**, 410–414 (1968). doi: 10.1145/363347.363368
48. Advanced Micro Devices Inc., FirePro S9150 Server GPU Datasheet (2014); https://usermanual.wiki/Document/AMDFirePROS9150DataSheet.2051023599.
49. A. Krizhevsky, I. Sutskever, G. E. Hinton, "Imagenet classification with deep convolutional neural networks" in *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, F. Pereira, C. J. C. Burges, L. Bottou, Eds. (Curran Associates, 2012).
50. D. C. Cireşan, U. Meier, L. M. Gambardella, J. Schmidhuber, Deep, big, simple neural nets for handwritten digit recognition. *Neural Comput.* **22**, 3207–3220 (2010). doi: 10.1162/NECO_a_00052; pmid: 20858131
51. R. Raina, A. Madhavan, A. Y. Ng, "Large-scale deep unsupervised learning using graphics processors" in *ICML '09: Proceedings of the 26th Annual International Conference on Machine Learning* (ACM, 2009), pp. 873–880.
52. N. P. Jouppi *et al*., "In-datacenter performance analysis of a tensor processing unit" in *ISCA '17: Proceedings of the 44th Annual International Symposium on Computer Architecture* (ACM, 2017).
53. D. Shapiro, NVIDIA DRIVE Xavier, world's most powerful SoC, brings dramatic new AI capabilities (2018); https://blogs.nvidia.com/blog/2018/01/07/drive-xavier-processor/.
54. B. W. Lampson, "Software components: Only the giants survive" in *Computer Systems: Theory, Technology, and Applications*, A. Herbert, K. S. Jones, Eds. (Springer, 2004), chap. 20, pp. 137–145.
55. D. L. Parnas, On the criteria to be used in decomposing systems into modules. *Commun. ACM* **15**, 1053–1058 (1972). doi: 10.1145/361598.361623
56. A. Danowitz, K. Kelley, J. Mao, J. P. Stevenson, M. Horowitz, CPU DB: Recording microprocessor history. *Queue* **10**, 10–27 (2012). doi: 10.1145/2181796.2181798
57. WikiChip LLC, WikiChip (2019); https://en.wikichip.org/.
58. B. Lopes, R. Auler, R. Azevedo, E. Borin, "ISA aging: A X86 case study," presented at WIVOSCA 2013: Seventh Annual Workshop on the Interaction amongst Virtualization, Operating Systems and Computer Architecture, Tel Aviv, Israel, 23 June 2013.
59. H. Khan, D. Hounshell, E. R. H. Fuchs, Science and research policy at the end of Moore's law. *Nat. Electron.* **1**, 14–21 (2018). doi: 10.1038/s41928-017-0005-9
60. J. Edmonds, R. M. Karp, Theoretical improvements in algorithmic efficiency for network flow problems. *J. Assoc. Comput. Mach.* **19**, 248–264 (1972). doi: 10.1145/321694.321699
61. D. D. Sleator, R. E. Tarjan, A data structure for dynamic trees. *JCSS* **26**, 362–391 (1983).
62. R. K. Ahuja, J. B. Orlin, R. E. Tarjan, Improved time bounds for the maximum flow problem. *SICOMP* **18**, 939–954 (1989). doi: 10.1137/0218065
63. A. V. Goldberg, S. Rao, Beyond the flow decomposition barrier. *J. Assoc. Comput. Mach.* **45**, 783–797 (1998). doi: 10.1145/290179.290181
64. T. B. Schardl, neboat/Moore: Initial release. Zenodo (2020); https://zenodo.org/record/3715525.

ACKNOWLEDGMENTS

We thank our many colleagues at MIT who engaged us in discussions regarding the end of Moore's law and, in particular, S. Devadas, J. Dennis, and Arvind. S. Amarasinghe inspired the matrix-multiplication example from the Software section. J. Kelner compiled a long history of maximum-flow algorithms that served as the basis for the study in the Algorithms section. We acknowledge our many colleagues who provided feedback on early drafts of this article: S. Aaronson, G. Blelloch, B. Colwell, B. Dally, J. Dean, P. Denning, J. Dongarra, J. Hennessy, J. Kepner, T. Mudge, Y. Patt, G. Lowney, L. Valiant, and M. Vardi. Thanks also to the anonymous referees, who provided excellent feedback and constructive criticism. **Funding:** This research was supported in part by NSF grants 1314547, 1452994Sanchez, and 1533644. **Competing interests:** J.S.E. is also employed at Nvidia, B.W.L. is also employed by Microsoft, and B.C.K. is now employed at Google. **Data and materials availability:** The data and code used in the paper have been archived at Zenodo (*64*).

10.1126/science.aam9744

RESEARCH ARTICLE SUMMARY

PLASMID EVOLUTION

Unexpected conservation and global transmission of agrobacterial virulence plasmids

Alexandra J. Weisberg, Edward W. Davis II, Javier Tabima, Michael S. Belcher, Marilyn Miller, Chih-Horng Kuo, Joyce E. Loper, Niklaus J. Grünwald, Melodie L. Putnam, Jeff H. Chang*

INTRODUCTION: Plasmids are autonomously replicating, nonessential DNA molecules that accelerate the evolution of many important bacterial-driven processes. For example, plasmids spread antibiotic resistance genes, which are a pressing problem for human and animal health. Plasmids can also encode complex traits that allow bacteria to interact intimately with eukaryotes. Acquisition of an oncogenic tumor-inducing (Ti) or root-inducing (Ri) plasmid by saprophytic soil agrobacteria changes them into pathogens capable of genetically transforming and causing disease in a broad range of plant species.

Plasmids are also biotechnology tools that can advance our understanding of life. They can be used to generate organisms with unusual traits and innovative applications. The potential for using oncogenic plasmids to accelerate research was recognized early in their discovery. Along with strains of agrobacteria, disarmed plasmids are mainstays as tools in plant biology and plant genetic engineering.

RATIONALE: Inferring evolutionary relationships is foundational for classifying plasmids, accurately assessing the influence of plasmids on disease outbreaks, developing appropriate strategies for mitigating disease, and expediting efforts to leverage plasmid diversity for biotechnology. However, such research is complicated because plasmids consist of diverse structural variants and are extraordinarily dynamic, modular molecules that can be reshuffled and broadly transmitted horizontally.

We focused on oncogenic plasmids of agrobacteria because of their important roles in causing disease and as biotechnology tools. Two genomic datasets were developed. One consisted of diverse, broadly sampled historical strains and was intended to serve as the basis for an evolutionary framework. The other consisted of contemporaneous strains hierarchically sampled from managed plant production sites, for the purpose of calibrating epidemiology methods. The datasets were combined to identify epidemiological patterns.

RESULTS: We combined analyses of chromosomal ancestry and plasmids to uncover their contributions and accurately model the global spread of disease. Phylogenetic, genomic, and time tree analyses of thousands of strains from the Rhizobiales order yielded a robust phylogenetic history of agrobacteria. We developed a strategy that uses phylogenetic and network approaches as well as different scales of genetic information to infer the evolution of diverse oncogenic plasmids. By combining results, we uncovered global epidemiological patterns supporting movement of pathogens clonally and plasmids horizontally in space and time.

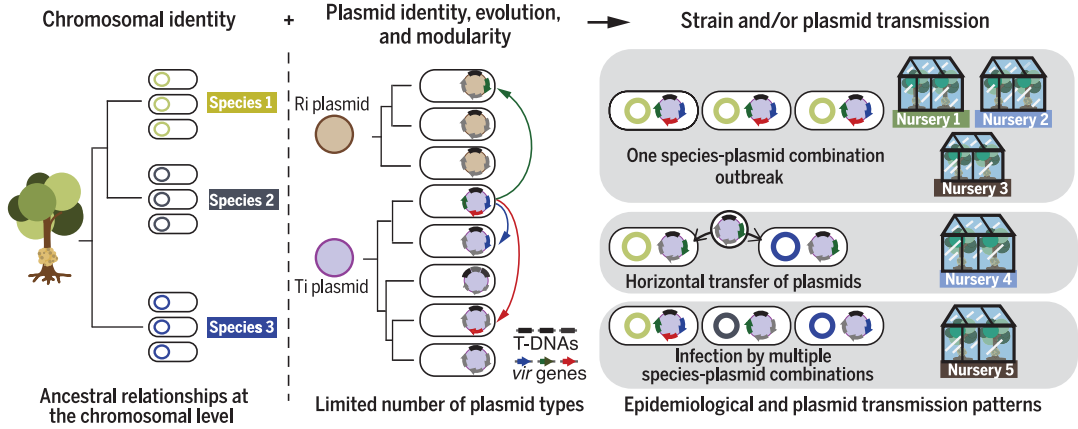
This study has three major findings: (i) Lineages of agrobacteria emerged independently and at different times from within a genus-level group that also circumscribes multiple lineages of rhizobia. (ii) Agrobacterial Ti and Ri plasmids are descended from only six and three lineages (types), respectively. Few evolutionary events are sufficient to explain the relationships observed among types. Each type is subject to different pressures and shows different degrees of within-group variation, but their evolution is nonetheless guided by similar principles. The extent of modularity is high, and genes and functional modules are frequently reshuffled via recombination within conserved loci. Yet plasmid diversification is nonetheless constrained by the spatial structure of loci that interact genetically. (iii) Transmission of oncogenic plasmids, especially within agricultural settings, promotes the massive spread of disease.

ON OUR WEBSITE

Read the full article at https://dx.doi.org/10.1126/science.aba5256

CONCLUSION: Our strategy for inferring the evolution and transmission of virulence plasmids has potential applications in plant and human or animal health and food safety, as well as for understanding the ecology and evolution of other plasmid-mediated processes such as mutualistic symbioses. In addition, this strategy can be applied to study other mobile and modular elements, such as integrative conjugative elements and pathogenicity or symbiosis islands. We have shown that plasmids once viewed as too diverse to be classified have distinct lineages, and that accurate modeling of the spread of disease can be accomplished by robustly defining their evolutionary relationships. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: changj@science.oregonstate.edu
Cite this article as A. J. Weisberg *et al.*, *Science* **368**, eaba5256 (2020). DOI: 10.1126/science.aba5256



Combined genomic analyses of chromosomal and plasmid identities to model disease spread. Genomic data from hundreds of strains of agrobacteria were parsed and analyzed to infer the evolutionary histories of chromosomes and oncogenic Ti and Ri plasmids. The data were overlaid to uncover the roles of bacteria and plasmids in the global spread of disease.

RESEARCH ARTICLE

PLASMID EVOLUTION

Unexpected conservation and global transmission of agrobacterial virulence plasmids

Alexandra J. Weisberg[1], Edward W. Davis II[1,2], Javier Tabima[1], Michael S. Belcher[1], Marilyn Miller[1], Chih-Horng Kuo[3], Joyce E. Loper[1,2,4], Niklaus J. Grünwald[4], Melodie L. Putnam[1], Jeff H. Chang[1,2,5]*

The accelerated evolution and spread of pathogens are threats to host species. Agrobacteria require an oncogenic Ti or Ri plasmid to transfer genes into plants and cause disease. We developed a strategy to characterize virulence plasmids and applied it to analyze hundreds of strains collected between 1927 and 2017, on six continents and from more than 50 host species. In consideration of prior evidence for prolific recombination, it was surprising that oncogenic plasmids are descended from a few conserved lineages. Characterization of a hierarchy of features that promote or constrain plasticity allowed inference of the evolutionary history across the plasmid lineages. We uncovered epidemiological patterns that highlight the importance of plasmid transmission in pathogen diversification as well as in long-term persistence and the global spread of disease.

Agricultural ecosystems promote the rapid evolution and diversification of pathogens (*1*). These ecosystems increase pathogen population sizes, lower barriers for transmission, and increase opportunities for horizontal exchange of virulence genes. Understanding the genetic basis for how pathogens emerge and diversify in agricultural ecosystems is foundational for determining their spread and assessing risks. Such knowledge is critical to policies for improving plant health and preparing against disease outbreaks to increase global food security (*2*).

Plasmids that confer virulence and antimicrobial resistance are evolutionary drivers of pathogenic bacteria that affect plant, human, and animal health (*3*–*6*). Some plasmids can mediate conjugation and transmit horizontally within and across species to diversify pathogen populations. The development of strategies to infer horizontal transfer of plasmids is crucial for accurately assessing disease outbreaks and instituting measures to limit risks from disease. However, it is difficult to trace dissemination of plasmids and integrate findings with analyses of chromosomes, because plasmids can be horizontally transferred and hence can defy inference of evolutionary histories (*7*). In addition, plasmids tend to have few conserved regions, high numbers of repeated sequences, high rates of gene exchange, and many structural variants. Even core genes such as those associated with replication or mobility are susceptible to recombination.

Agrobacteria are a diverse, polyphyletic group, and its members are common in soils and on many species of plants. These bacteria have multipartite genomes with two chromosomes and nonvirulence plasmids, which are extremely diverse among agrobacteria (*8*). Pathogenic strains additionally carry conjugative oncogenic tumor-inducing (Ti) or root-inducing (Ri) plasmids and can infect plants to cause crown gall or hairy root diseases, respectively (*9*). These diseases are incurable and persist for the duration of the lives of infected plants, which continually release opines, metabolites that are central to the ecology and epidemiology of agrobacteria.

Oncogenic plasmids are exceptionally well characterized because of their applications in plant biotechnology and hence are models for understanding the impact of plasmids on disease ecology (*10*, *11*). Core to oncogenic plasmids are *repABC* replication genes, *tra* and *trb* interbacterial conjugation genes, and *vir* genes. Vir proteins are necessary to process and escort a region, the transfer DNA (T-DNA), of oncogenic plasmids into host cells, where it recombines into the genome to genetically transform host plants. T-DNAs include oncogenes that reprogram transformed cells to proliferate. One oncogene present on T-DNAs in all oncogenic plasmids is *tms1* (*aux1/iaaM*), which is involved in the synthesis of auxins, a class of plant growth–promoting hormones (*12*). T-DNAs also include genes responsible for the synthesis of opines. More than 20 chemically diverse opines have been identified among oncogenic plasmids (*13*). Opines are nutrients for the pathogen and act as signals that trigger replication and interbacterial conjugation of oncogenic plasmids. Genes necessary for uptake and catabolism of opines are spatially separated from cognate synthesis genes and are located outside of T-DNAs. This arrangement led to the hypothesis that instigating pathogens are privileged in accessing the opines (*14*).

Beyond the core set of genes, oncogenic plasmids vary extensively in composition and structure (*15*). This diversity is a major challenge, best expressed by the sentiment that it is practically impossible to reconstruct the evolution of agrobacteria and oncogenic plasmids, even with the availability of an extraordinarily large genomic dataset (*16*). Here, we overcame this challenge and analyzed genomic data from hundreds of strains, and applied the results to infer the evolutionary history and transmission of oncogenic plasmids.

Chromosomal ancestry

A robust phylogeny of bacteria is an essential foundation for understanding the evolution and transmission of the plasmids that they host. Hence, it was critical to reconcile controversies over the complex evolutionary relationships among agrobacteria (see data S1 for synonymous terms) (*17*, *18*). Sequenced strains were reclassified in the context of a dataset of nearly 1500 additional strains that represent families within the order Rhizobiales (data S1 and S2). Combined phylogenetic and genomic analyses indicated that agrobacteria, *Rhizobium*, *Allorhizobium*, *Neorhizobium*, *Ensifer/Sinorhizobium*, *Pararhizobium*, and *Shinella* are related at a level of a genus, referred to as the agrobacteria-rhizobia complex (ARC; fig. S1 and data S3). Within the ARC, groups traditionally known as *Agrobacterium tumefaciens*, *Agrobacterium rhizogenes*, and *Agrobacterium vitis* are called biovar (BV) 1, BV2, and BV3, respectively (*19*). BV1 was previously subclassified into genomospecies (*20*). Whole-genome analysis suggested that most genomospecies (excepting G7 and G8, which each represent three groups) are commensurate with species-level groups (data S4). BV2 is homogeneous and represents a single species-level group. BV2 is sister to a newly defined clade of pathogenic agrobacteria (BV2-like). BV3 forms multiple species-level groups.

A single lineage of agrobacteria was previously estimated to have diverged from rhizobia between ~250 million and ~150 million years ago (*21*, *22*). Considering the new interpretation of their phylogeny, we reevaluated estimates of divergence times (Fig. 1, fig. S2, and data S5). Lineages of agrobacteria emerged independently, at different times in the history of the ARC, and are more closely related to lineages of rhizobia than to each other. BV1 emerged 48 ± 10 million years ago, and its clade has substructure with long branch lengths, which suggests that species-level groups diverged early in its history. The BV2 lineage is

[1]Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331, USA. [2]Molecular and Cellular Biology Program, Oregon State University, Corvallis, OR 97331, USA. [3]Institute of Plant and Microbial Biology, Academia Sinica, Taipei 11529, Taiwan. [4]Horticultural Crops Research Laboratory, USDA Agricultural Research Service, Corvallis, OR 97331, USA. [5]Center for Genome Research and Biocomputing (CGRB), Oregon State University, Corvallis, OR 97331, USA.
*Corresponding author. Email: changj@science.oregonstate.edu

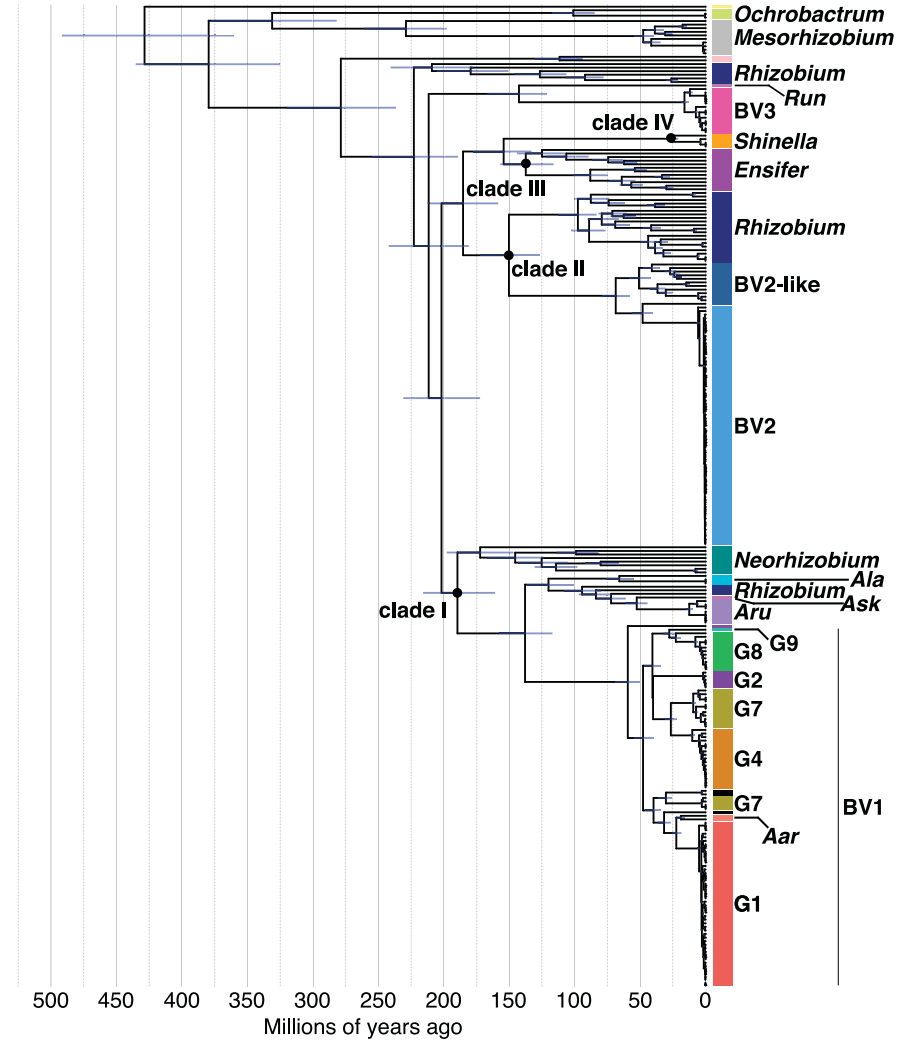


Fig. 1. Time-calibrated phylogenetic tree of the agrobacteria-rhizobia complex. Blue horizontal bars indicate confidence intervals for each split. Clades I to IV are defined in fig. S1. Key groups are labeled; three-letter codes are for select species of rhizobia or agrobacteria without biovar classifications: *Run* (*R. undicola*), *Ala* (*A. larrymoorei*), *Ask* (*A. skierniewicense*), *Aru* (*A. rubi*), and *Aar* (*A. arsenijevicii*); synonyms are listed in data S1. Each strain is colored according to its species-level classification. Two groups (black bars) within BV1 are not associated to established genomospecies.

estimated to have emerged only 1.6 ± 0.5 million years ago, and its short branch lengths and low genetic diversity are consistent with the possibility that BV2 is recovering from a recent bottleneck (fig. S2). In the time tree, BV3 diverged independently from other lineages of agrobacteria (Fig. 1). However, its relationships in the time tree are incongruent with those of the phylogenetic tree (fig. S1).

Plasmid identities

We next addressed the difficult task of resolving the relationships of the oncogenic plasmids. Because plasmid genes do not meet the assumptions of traditional phylogenetic methods, we used multiple approaches to compare and cross-validate findings. A total of 143 oncogenic plasmid sequences were identified. We analyzed different levels of genetic features, including signatures of subsequences of length k (k-mers), a 43-core gene phylogeny, a phylogenetic network derived from 144 single-copy genes present in at least 40% of all plasmids, and patterns of composition as well as organization. To our surprise, the results were consistent and allowed us to categorize the molecules into just nine distinct lineages of six Ti (type I–VI) and three Ri (type I–III) oncogenic plasmids (Figs. 2 and 3, figs. S3 to S9, and data S6 and S7). Type I and type IV Ti plasmids were further divided into two and three subtypes, respectively.

In the k-mer network, most types segregated into distinct graphs (Fig. 2A). However, type IV.c Ti plasmids are high-degree nodes that connect type I Ti plasmids to the other subtypes of type IV Ti plasmids. In the phylogenetic network, there were many splits, yet plasmids clustered via short edges into monophyletic clades (Fig. 2B). This topology is consistent with extensive and ancient recombination events prior to divergence of plasmid lineages. The two networks show that the type I Ti plasmids are most closely related to type IV plasmids, which is consistent with their similarities in structure and gene composition (fig. S10 and data S8). Other relationships were also observed in the phylogenetic network. Type I Ti plasmids are related to the type VI Ti plasmid (Fig. 2B). Type II plasmids are nested within the type III Ti plasmids and are more closely related to each other than to other types of Ti plasmids. Type V Ti plasmids are the most distinct of the sequenced Ti plasmids. Type II and type III Ri plasmids are more closely related to each other than to type I Ri plasmids.

Within this sequenced dataset, there are relationships among plasmids, bacterial species, and plant hosts, some of which had not been previously noted and could potentially be of use in biotechnology applications of agrobacteria (Fig. 2C and fig. S3). BV1 has the most diverse spectrum of oncogenic plasmids, with four types of Ti plasmids and type I Ri plasmids. Type III Ti plasmids are exclusively in BV1. BV2 tends toward having only one of the two minor variants of type 1.a Ti plasmids (fig. S15A). However, some have type II and type IV.c Ti plasmids as well as type III Ri plasmids. BV3 strains exclusively carry type IV.a, type IV.b, and type V Ti plasmids. Strains carrying type I Ti plasmids were isolated predominately from woody plants, whereas those with type III Ti plasmids were exclusively from herbaceous plants. With the exception of only two strains, those belonging to BV2 were cultured from woody plants. BV3 strains were exclusively cultured from grapevine, which is also host to BV1 and BV2 strains.

There are different degrees of variation within each type of oncogenic plasmid. We combined gene synteny, gene annotation, and sequence data for a comprehensive comparison of patterns of variation in plasmid types (Fig. 3A, figs. S5 to S9, and data S9). Type I.a and type II Ti plasmids are relatively conserved in gene composition and structure. Type I.b and type III Ti plasmids are more diverse, with gene presence/absence variation present in T-DNAs, opine-associated loci, and regions flanking the *vir* loci. Despite the lower sampling depth, a range in gene presence/absence was also observed among the Ri plasmids. Across all but type II and type V Ti plasmids, higher variation in gene composition is present within T-DNAs and proximal to the right border. Variation extends to the region neighboring the right border of T-DNAs in type I.b and type III Ti plasmids and all types

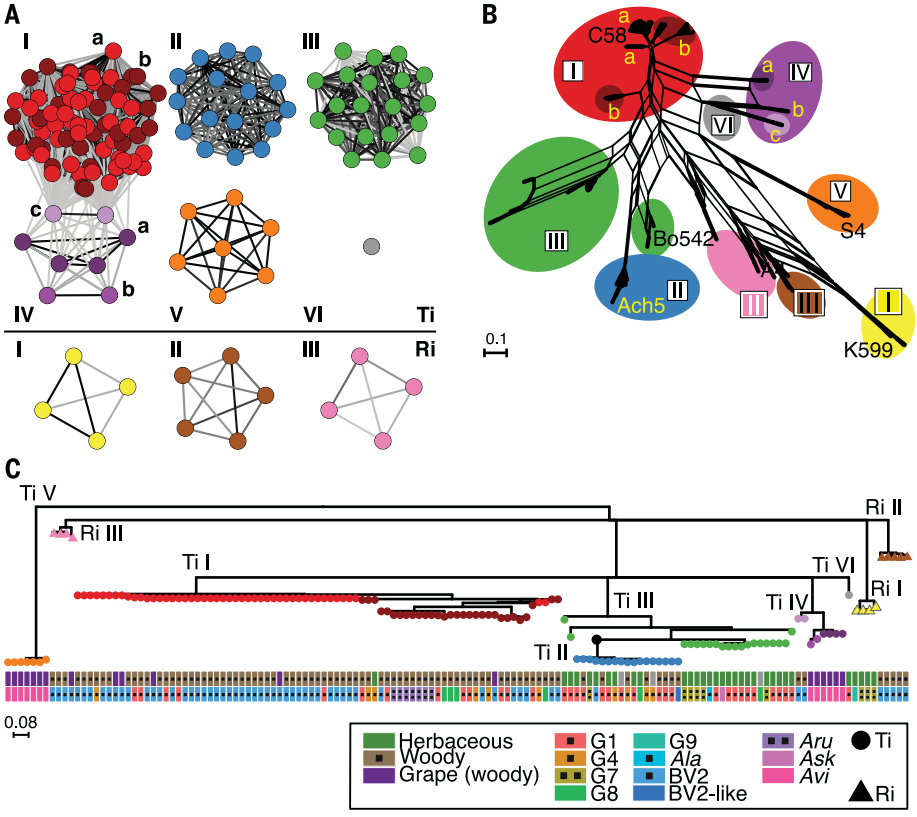


Fig. 2. Nine distinct lineages of oncogenic plasmid types. (**A**) Weighted undirected network of oncogenic plasmids. Nodes represent individual oncogenic plasmids and are colored according to the type and subtype of Ti (top) and Ri (bottom) plasmids. Darker edges indicate greater Jaccard similarity of *k*-mer signatures. (**B**) A split network of the oncogenic plasmids. Branch thickness indicates relative support for the split. Key reference strain plasmids are indicated. (**C**) Maximum likelihood tree constructed on the basis of concatenated sequences from 43 single-copy core genes (data S7). Tips (circles and triangles) are color-coded according to plasmid type. Colored panels in the top row below the tree denote the type of plant from which strains were cultured. Colored panels in the bottom row indicate the classification of the strains. The same color scheme for plasmid types is used in each of the three panels. The tree is midpoint-rooted. See fig. S3 for a more detailed and larger version of (C).

of Ri plasmids. In these plasmids, the region encodes proteins necessary for transport and catabolism of opines. As previously noted, noncore genes of most plasmid types are highly variable. The *tra* and *trb* loci of type I–III Ti plasmids have extensive small-scale changes.

Plasmid evolution

The range in variation across types of oncogenic plasmids suggests that each of the types has been shaped by different evolutionary processes. We modeled plasmid histories to understand these processes and infer origins and relationships. In doing so, a constant theme emerged: Core genes used to characterize oncogenic plasmids not only have low phylogenetic value but are in fact the major contributors to their variation. Their conservation among plasmids provides regions for recombination and reshuffling of genes to occur. We therefore analyzed plasmid types at different scales to infer patterns of evolution. Features of oncogenic plasmids were also accounted for separately because of their different evolutionary histories.

Different levels of modularity and variation in gene analogs contribute to oncogenic plasmids maintaining their function while adopting diverse gene composition and structure. The *vir* locus is central to virulence of agrobacteria yet is prone to exchange among plasmid types. Results from *k*-mer and phylogenetic analyses were not consistent with grouping on the basis of locus structure or with relationships among plasmid types (Fig. 4A and figs. S11 and S12A). The *vir* loci of type I.a, type I.b, type II, type III, and type VI Ti plasmids as well as type II and type III Ri plasmids share a recent common ancestor and follow one of the paths in the *k*-mer graph. Those of type IV and type V Ti plasmids and type I Ri plasmids share a recent common ancestor and follow the second path.

The *vir* locus is often acquired as separate modules. There are two *vir* loci in type IV.c Ti plasmids. The primary *vir* locus is hypothesized to have been acquired from a Ri plasmid. By itself, this locus is likely not functional because of multiple transposase genes interrupting *virA*, which encodes a regulator of the *vir* genes (fig. S12B). The second *vir* locus is a remnant that encompasses only *tzs*, *virA*, *virB1*, *virB2*, and *virB3* and has overall identity of 93% to the *vir* locus of type I Ti plasmids. We speculate that the *virA* allele of this fragment is necessary to complement the disrupted allele of the primary *vir* locus. In addition, *virE1-2* and *GALLS* genes are interchangeable submodules that are frequently inherited separately from the rest of the *vir* locus. This was observed for *virE1-2* genes among type IV–VI Ti and type II Ri plasmids (fig. S12). Likewise, the *vir* locus of type III Ri plasmids, despite having a common ancestor with that of type I Ti plasmids, carries *GALLS* instead of *virE1-2*.

T-DNAs are extraordinarily diverse in size as well as composition and are highly recombinogenic (Fig. 4B, fig. S13, and data S6). T-DNA transfer is largely dependent on a short right border sequence and a flanking *overdrive* enhancer sequence (fig. S14) (*23*–*25*). Border sequences are practically invariant among 213 T-DNAs, and this strict conservation of short sequences is a low barrier for generating alternative T-DNA–*vir* combinations and multiplexing T-DNAs in plasmids (fig. S14, A and B, and data S6). Diversification can be driven by gene gain and loss from T-DNAs with little to no consequence to the transformation process. Chimerization is another frequently permitted mechanism of diversification (figs. S15 and S16). The original T-DNA of the type I.a Ti plasmid (e.g., pTiC58) was extended, from left to right, by invasion of two additional T-DNAs. The last of the T-DNAs is one of two prominent variants that invaded and swept through the type I–IV Ti plasmids, potentially because of a selective advantage over others. This is the path from *acs* through the *6b* gene to the right border that cuts prominently across the gene synteny graph. The type III Ti plasmid has two left borders in T-DNA-1 because a second T-DNA displaced all but the most left-flanking *acs* gene of the original T-DNA. T-DNA-2 of the type VI Ti plasmid has three right border sequences and two sets of homologous, but nonparalogous, oncogenes. This T-DNA follows a complex path in the graph.

However, within this dataset, there is little evidence for exchange of T-DNAs between classes of oncogenic plasmids. With the exception of T-DNA-2 of type III Ri plasmids, T-DNAs of the Ti and Ri classes have distinct gene compositions and segregate into different regions of the gene synteny graph prior

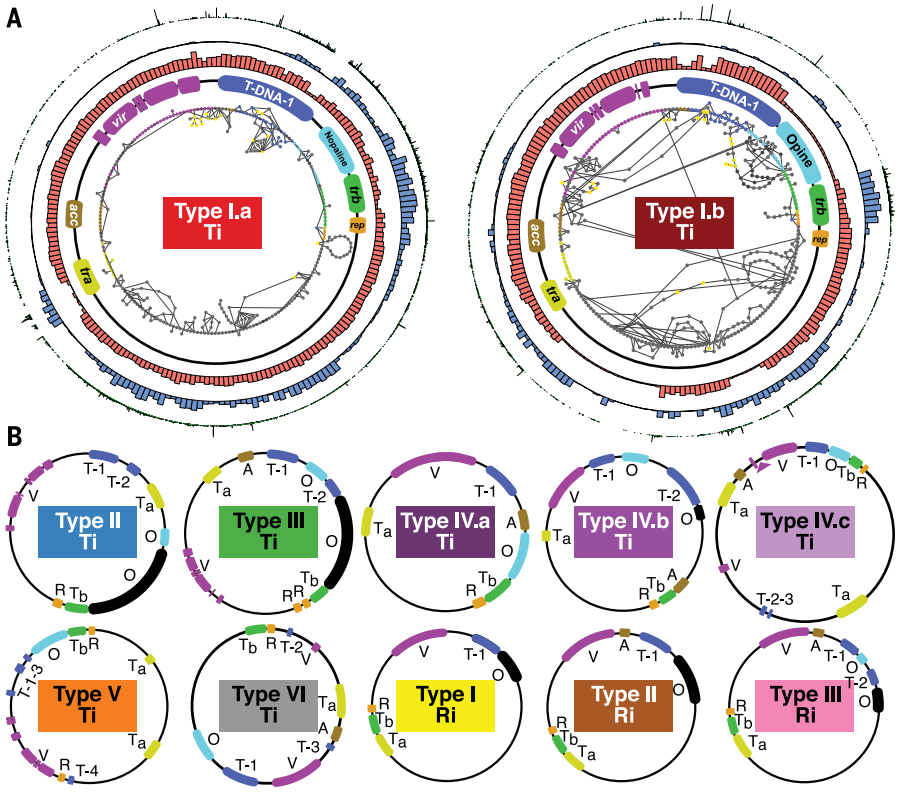


Fig. 3. Variations within and between types and subtypes of oncogenic plasmids. (**A**) Visualization of variation within plasmid subtypes. Circles, starting from the innermost circle, are a gene synteny graph, a plasmid map, and bar graphs representing relative depth of coverage of sequencing reads (sliding window of 1 kb), number of SNPs (sliding window of 1 kb), and number of soft-clipped reads (≥5 clipped reads). Nodes in the synteny graph represent genes present in ≥1 plasmid of a type. Nodes are connected if genes are adjacent in at least one plasmid. The network cycle represents the most common order of genes within a type. Different structures, such as gene presence/absence variation, rearrangements, or inversions present in plasmids, are shown as alternative paths in the cycle. Nodes are colored according to key functions and aligned to corresponding features in the plasmid map, or colored yellow for transposase- and insertion sequence–encoding genes. Soft-clipped reads are those that must be trimmed in order to align to the reference and can be evidence for a recombination breakpoint. Type I.a and type I.b Ti plasmids are presented as examples. (**B**) General structure and organization of type II–VI Ti plasmids, as well as their subtypes, and type I–III Ri plasmids. Simplified maps show locations and spatial relationships of key modules. They are colored and labeled with a letter code: A, *acc*; O, opine genes; R, *repABC*; T-1 to T-4, T-DNA 1–4; T_a, *tra*; T_b, *trb*; V, *vir* genes. The triangle represents an insertion sequence. Visualizations are as shown in (A) for each of the plasmid types and subtypes (figs. S5 to S9).

to converging on a common subset of opine synthesis genes (Fig. 4B).

The structural arrangement of modules affects the degree of flexibility of oncogenic plasmid types. Most T-DNAs have a gene for synthesizing an opine; opines are conjugates between a keto acid and an amino acid (K-A), a sugar and an amino acid (S-A), or two sugars (S-S). All oncogenic plasmids have at least one T-DNA with an opine synthesis gene, and plasmid classification traditionally relied on the opine synthesized, despite assumptions of frequent exchange of opine genes (Fig. 4). Although variation of opine genes does exist, there is no evidence for the generalization of rampant exchange among all types of plasmids (Fig. 4C). Opine variation is influenced by the proximity of genes to the right borders of T-DNAs. When the cognate opine anabolism and catabolism/uptake genes are closely linked and separated by only the right border, swapping of opine loci and the border sequence can occur with little constraint (figs. S5, S6, S14, S17, and S18). Type I.b Ti plasmids are very promiscuous (fig. S5). We hypothesize that swapping is mediated by a process that involves nonhomologous recombination within oncogenes. In type III Ti plasmids, exchange is predicted to be mediated by homologous recombination among oncogenes (fig. S6). Because K-A and S-A genes are associated with different sets of oncogenes, swapping is most frequently limited to within their structural groups of opines (fig. S17 and data S6). Type I.a Ti plasmids are structurally similar to type I.b Ti plasmids and have the potential to diversify opines but are far more limited. We suggest that the low diversity is a consequence of the ancestral type I.a Ti plasmid experiencing a bottleneck along with BV2 (Fig. 1).

In other types of plasmids, important functional modules separate opine genes, and multiple recombination events would need to occur to acquire the two interdependent modules. For example, opine genes of type II plasmids are interrupted by *tra* genes (fig. S6). In type IV.a Ti plasmids, genes necessary for anabolism and catabolism of nopaline are separated by those involved in the catabolism of a second opine, agrocinopine (fig. S7). The *acs* gene, cognate to *acc* and necessary for synthesizing agrocinopine, is within and adjacent to the left border of the T-DNA (Fig. 4B and fig. S13). This organization is restrictive; for opine swapping to occur, either a larger fragment that included most of the T-DNA would also need to be exchanged, or the two modules for nopaline would need to be acquired separately.

The *trb* locus is associated with a high number of single-nucleotide polymorphisms (SNPs) and has diverse alleles (figs. S3 to S9 and S19). These data are consistent with the possibility of recombination within the *trb* locus, and because *trb* is adjacent to opine loci, it likely contributes to the swapping of opine cassettes in many types of Ti plasmids. Some type III Ti plasmids have additional copies of *trb* genes located distal to the main *trb* locus, likely due to duplication resulting from recombination within the locus (fig. S18B). In contrast, in the Ri class, *trb* is paired with *tra* and is separated from opine loci by large regions with no genes involved in virulence or plasmid maintenance (fig. S9). This difference in spatial relationship between classes likely constrains recombination between Ti and Ri plasmids. However, *tra* and *trb* could mediate recombination with homologous loci of the diverse non-oncogenic plasmids that are prevalent in the ARC (fig. S20).

All findings were synthesized to model the evolution of oncogenic plasmids (Fig. 5). Ti and Ri plasmids are distantly related and are hypothesized to have emerged from an ancestral proto-oncogenic plasmid (*13*, *26*). We hypothesize that this ancient replicon carried *vir*, *tzs*, *acs*, and an opine synthesis gene flanked by T-DNA borders (fig. S21 and data S8). The *tzs* gene is a paralog of *ipt* (*tmr*) that is distal to the T-DNA and not transferred into plant cells (*27*). The *tzs* and *ipt* genes encode isopentenyltransferases, enzymes that catalyze the first step in the synthesis of the plant growth–promoting hormone cytokinin (*28*). However, cytokinins derived from Tzs have been implicated in regulating *vir* gene

expression and promoting virulence (*29*, *30*). The proto-oncogenic plasmid was suggested to have originated as a catabolic plasmid. The largest set of genes inferred to have a shared evolutionary history includes 52 genes (36% of those analyzed), of which 10 are implicated in the uptake of nutrients or metabolism and another 18 have unknown functions (data S8 and figs. S22 and S23). This set also includes *repABC*. We propose that the proto-oncogenic plasmid led to either the Ti or Ri class of plasmid and recombined with a separate *repABC* plasmid to yield the other class.

The next innovation was the acquisition of oncogenes within T-DNA border sequences. Every one of the sequenced Ti plasmids has at least one T-DNA that circumscribes *tms1* and *tms2* (*aux2/iaaH*), and all Ri plasmids have at least one T-DNA with *tms1* (Fig. 4B and data S6). The *tms2* gene cooperates with *tms1* in auxin biosynthesis (*12*). Data are consistent with the ancestors of Ti and Ri plasmids independently acquiring *tms1* from different nonagrobacterial sources (fig. S24). Moreover, even though *tms1* and *tms2* are a functional module and are consistently linked in T-DNAs of Ti plasmids, *tms2* was potentially acquired separately from *tms1* by an ancestral Ti plasmid (fig. S24C). The evidence suggests that in Ti plasmids, *ipt* was derived from a duplication of *tzs* and the paralog was incorporated within the T-DNA region (fig. S21).

An unexpectedly small number of events are sufficient to relate plasmid lineages (Fig. 5 and figs. S3, S4, and S12 to S24). Single-copy genes of oncogenic plasmids were grouped into just 18 sets inferred to have similar evolutionary histories (fig. S23). Five sets were sufficient to represent 80% of the genes analyzed. In contrast, *tra* and *trb* genes are distributed across 50% of the sets. Type IV.c Ti plasmids are cointegrates of a type I Ti plasmid and a Ri plasmid. The two sequenced type IV.c Ti plasmids have edges to every type I.a node, are nearly twice the size of pTiC58, and have a second *tra* locus located between two *vir* loci (Fig. 1A and fig. S10). In a modification to a previously proposed model, we suggest that all subtypes of the type IV plasmids are derived from the same lineage of the co-integrated ancestor and that type IV.a and type IV.b Ti plasmids are streamlined variants (*15*). The type VI Ti plasmid is also likely a cointegrate. This plasmid is novel in that it horizontally acquired, via T-DNA invasion, a second set of *tms1* and *tms2*, which is unlike the first set found in all types of Ti plasmids (fig. S24). Type II Ti plasmids are hypothesized to have recently emerged from a rearrangement event within the type III Ti plasmids. Type V Ti plasmids are more closely related to Ri plasmids and are the most genetically and structurally distinct of the Ti plasmids. The multiple changes and mosaic structure make its evolutionary history difficult to interpret. Its unique structure is predicted to limit opportunities for productive recombination events with other oncogenic plasmids. Although Ti and Ri classes are very distantly related, they can co-reside in cells and exchange regions. Strain Di1411, for example, carries a type I.a Ti plasmid and a type II Ri plasmid.

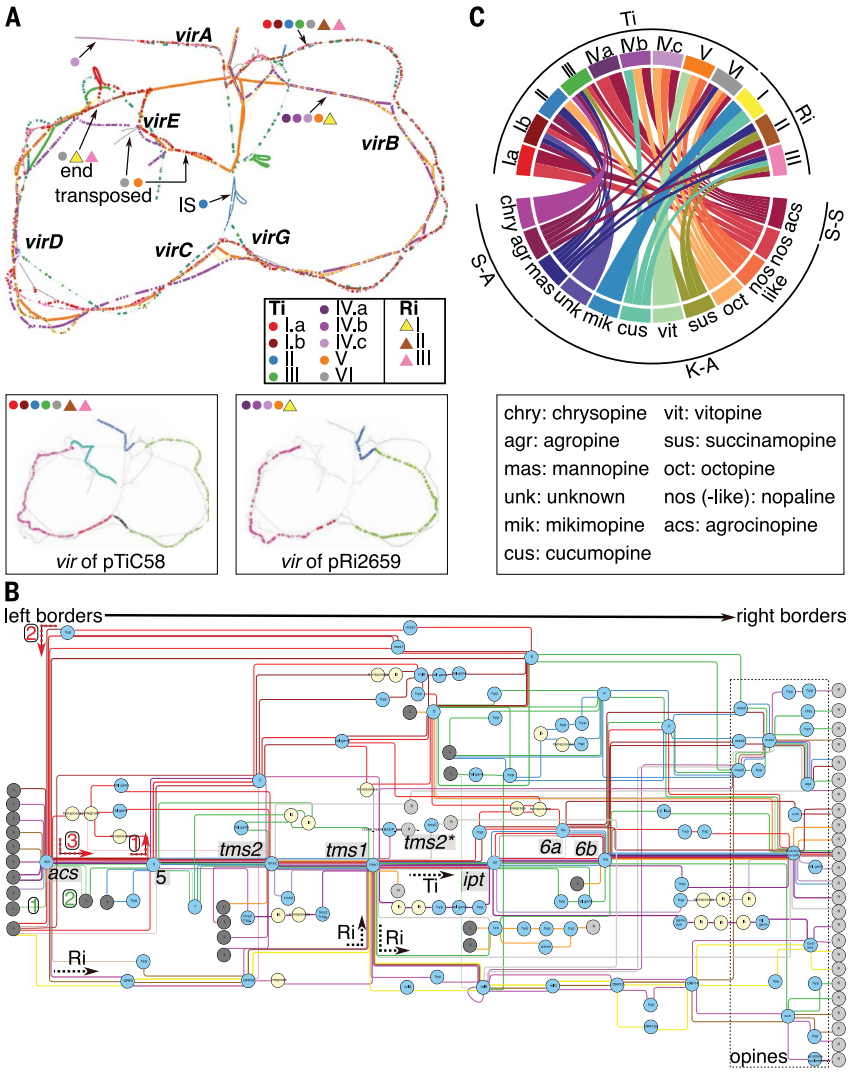


Fig. 4. Diversity of virulence loci of oncogenic plasmids. (**A**) Compacted de Bruijn graph constructed from component *k*-mers of sequenced *vir* loci. Traces are colored according to plasmid type and subtype. Key variations and *vir* gene regions are indicated; IS, insertion sequence. Large regions that interrupt the *vir* loci were not included. Panels at bottom show the traces of pTiC58 and pRi2659, which exemplify the two main traces. Traces in the inset are colored to differentiate between gene loci. (**B**) Gene synteny graph of 213 T-DNAs. The main Ti and Ri paths are labeled with dotted arrows. Paths of the three T-DNAs that form the chimera of the type I.a Ti plasmids are indicated by numbers in red. The two left borders of the chimeric T-DNA-1 of the type III Ti plasmids are labeled with numbers in green. Nodes are colored according to category (dark gray, left border; light gray, right border; blue, gene; yellow, insertion sequence). The *tms2** gene is an independently acquired homolog that failed to cluster with *tms2*. Edges are colored according to plasmid type. Line weight is normalized to the proportion of plasmids within a given type. (**C**) Circos plot relating oncogenic plasmids to synthesized opine variants (S-A, conjugates of sugars and amino acids; K-A, conjugates of keto acids and amino acids; S-S, conjugates of sugars). See figs. S11 and S13 for larger variations of (A) and (B).

Modeling disease spread

We next used this evolutionary framework to identify epidemiological patterns among strains collected across the world and spanning nearly a century of collection. The framework allowed us to analyze strains and plasmids as independently transmitted entities to accurately model disease spread. An additional 66 genome sequences of hierarchically sampled strains were used to guide grouping of the agrobacterial strains and oncogenic plasmids

to ranks below species and subtypes, respectively (data S10). Strains of G1, G4, G7 (three species-level groups), and G8 (three species-level groups) of BV1 as well as BV2 were divided into 124 unique genotypes. Type I–III Ti plasmids were divided into 40 distinct plasmid clusters (data S11 to S19).

Multiple patterns underscore aspects of agricultural ecosystems that promote the diversification and spread of pathogens (Fig. 6 and fig. S25). Nonpathogenic, plasmid-lacking strains were isolated from healthy and diseased individuals (data S12, S15, S16, and S17). For example, three nonpathogenic strains were cultured from symptomatic plants at facility C7_N18, a location that also had pathogenic strains, differing by ~130,000 to ~290,000 SNPs (Fig. 6A and data S17). As potential recipients of virulence plasmids, nonpathogenic strains, which are common in soils and in association with plants, represent a standing pool of genetic diversity for the evolution of new pathogen genotypes. Strain-plasmid combinations can persist over the long term within agricultural ecosystems. Strains LMG 267 (type II Ti) and B140/95 (type II Ti) were collected 60 years apart yet are members of the same genotype and plasmid cluster, with 7 and 0 SNP differences, respectively (Fig. 6B and fig. S25B). Pathogens can be transmitted between agricultural and unmanaged ecosystems. Three members of a BV2 genotype (type I.a Ti) linked a natural ecosystem, an agricultural ecosystem, and an undocumented location in a different country (Fig. 6C and fig. S25C). The three strains and their Ti plasmids have ≤5 and 0 SNP differences, respectively, among them. S6_N30 and S6_N25 are almost 240 km apart, but two other plant production facilities are located within 1.5 km of S6_N30 and could have bridged the link to S6_N25.

Agricultural ecosystems can have recurrent infections. Locations in which different genotype-plasmid combinations were detected had likely experienced independent infections. Facility S2_N7 had eight strains representing seven different genotype-plasmid combinations (Fig. 6D and fig. S25, I to M). Over a 12-year span, S8_N32 was infected by G1 (type III Ti) strains that have >50,000 and >350 SNP differences among strains and plasmids, respectively (fig. S25, N to P). An agricultural ecosystem can also have disease reservoirs; this idea is supported by the presence of multiple strains of the same genotype-plasmid combination on different individuals. Location C7_N18 had 27 strains of the same G7 genotype (type III Ti) on four individuals of three different host species (Fig. 6A and fig. S25Q).

There were at least seven cases in which global distribution of plants is hypothesized to have contributed to the transmission of a strain-plasmid combination. These are patterns in which genotype and plasmid nodes are connected by multiple edges. Patterns by themselves are insufficient for differentiating between direct and indirect transmission routes. However, one case is highlighted that includes a facility that produces plants for wholesalers and could be a common source. Strains belonging to a G1 genotype (type III Ti) were identified from facility C4_N13 (Fig. 6E and fig. S25E). Strains of the same genotype-plasmid combination were later identified in two other facilities. The strains and plasmids from these three locations have ≤10 and 0 SNP differences, respectively, among them. The dataset has other instances in which sets of closely related strains marginally exceeding the >15 SNP difference threshold have plasmids that belong to the same cluster. It is thus possible that there were more strain-plasmid transmissions than reported.

Horizontal transmission of plasmids greatly diversifies and amplifies the spread of pathogens. Fifteen networks have a plasmid cluster

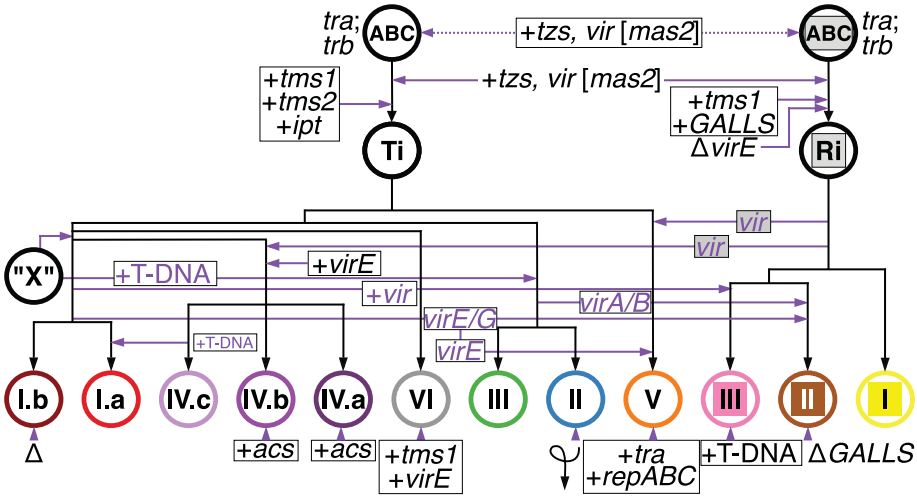


Fig. 5. Model of the evolution of oncogenic Ti and Ri plasmids. Genes in boxes were acquired from unknown sources. The cluster of *tzs*, *vir* [*mas*] genes is predicted to have been acquired once (dotted arrows) and then transferred from one plasmid backbone to the other (solid arrows). Genes in purple were acquired horizontally from the indicated sources. Purple arrows represent major horizontal acquisition events. Circle with "X" depicts an undefined plasmid hypothesized to be the donor of the prominent T-DNAs that swept through the type I–IV Ti plasmids.

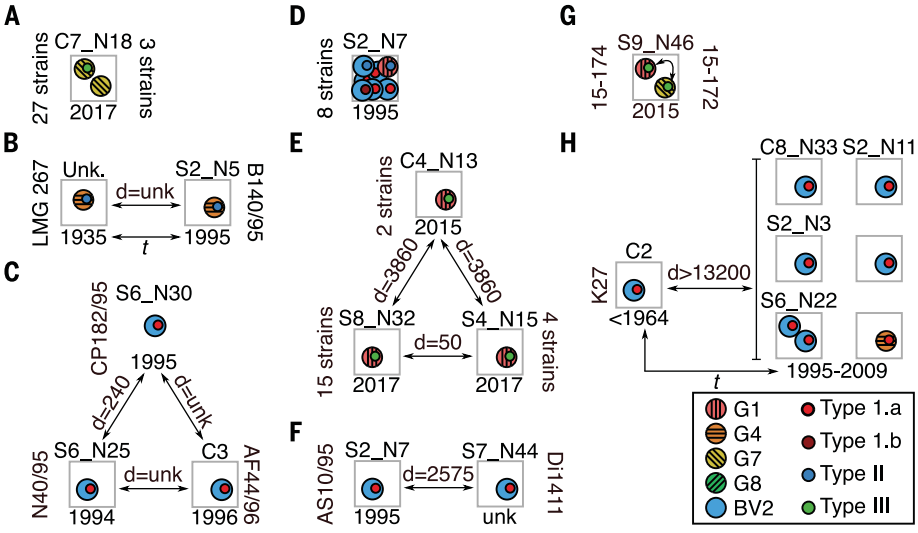


Fig. 6. Spatiotemporal transmission of strains and plasmids. (**A** to **H**) Patterns revealed in undirected networks combining strain genotypes and Ti plasmid clusters (fig. S25). Key: gray square, agricultural ecosystem; large circle, species-level group; small circle, type or subtype of Ti plasmid. The location is labeled with a coded identifier (top); strain name(s) (left or right side); date isolated (bottom; unk, unknown). Double-headed arrows link locations and show approximate distance (*d*, kilometers) and/or time (*t*, years). The large circle without a small circle in (A) represents a nonpathogenic strain that lacks an oncogenic plasmid.

as a central hub connected to multiple genotype nodes (fig. S25). We use the simple case of the type I.a Ti plasmid of strain Di1411 as an example (Fig. 6F and fig. S25F). This plasmid has only two SNP differences relative to the type I.a Ti plasmid of another BV2 strain. In contrast, the two strains linked by the type I.a Ti plasmid cluster have more than 22,000 SNP differences in their chromosomes. The most pronounced example has 28 BV2 strains and one G1 strain collected from more than 20 locations that have type I.a Ti plasmids with no more than 13 SNP differences among them (fig. S25R and data S11). The clustering of these 29 plasmids is not a consequence of low SNP diversity. The total of 49 type I.a Ti plasmids in this study formed 10 unique clusters that vary by nearly 5000 SNP differences. The diversity of BV2 strains is also not a consequence of the conservative threshold used to define genotypes. Even at 1000 SNP differences, the BV2 strains would separate into 17 different genotypes.

Management practices in agricultural ecosystems have been suggested to increase opportunities for plasmid conjugation (*31*). Supporting evidence is found in three networks in which the same location is mapped to a plasmid node and multiple associated genotype nodes. Strains of a G1 and G7 genotype carrying type III Ti plasmids with 0 SNP differences were identified in facility S9_N46 (Fig. 6G and fig. S25G). In the second example, one strain of a BV2 genotype and three strains of a G1 genotype were collected from facility S2_N7 (fig. S25K). Their type II Ti plasmids have 0 SNP differences among them. Last, a strain of a G4 genotype and two strains of closely related BV2 genotypes have type I.a Ti plasmids with ≤1 SNP difference (fig. S25S).

Plasmid dissemination can also lead to the temporal spread and persistence of disease. The type I.a Ti plasmid of strain K27, collected prior to 1964, has ≤2 SNP differences relative to seven other type I.a Ti plasmids present in strains collected during the period 1995–2009 (Fig. 6H and fig. S25H). As previously stated, this is likely not a reflection of low SNP diversity within the type I.a Ti plasmids.

We found one instance in which two closely related strains have distantly related Ti plasmids, suggesting independent acquisition events. Strain Z4/95 and strain K27 differ by only 48 SNPs (fig. S25, L and T, and data S19). The type I.a Ti plasmid of Z4/95 has ~2400 SNP differences in comparison to the type I.a plasmid of K27 and does not cluster with any other type I.a Ti plasmid (data S11).

Discussion

Modularity confers phenotypic robustness and allows oncogenic plasmids to diversify. But diversification is opposed by selective forces that preserve genetic and physical links necessary to maintain functionality of the plasmids. By accounting for variation and constraints, we were able to sort oncogenic plasmids into defined lineages and infer their evolution. We revealed the most commonly observed types, which are foundational for understanding types yet to be discovered.

Diversification of pathogen populations by plasmid transmission is promoted by the agricultural industry, where large populations of susceptible host species are intensively managed and often clonally propagated within the same facilities, individuals are moved across the world, and disease can persist undetected for extended periods of time. Although clinical settings have some similar features that promote diversification, antibiotic use is suggested to result in a recurring pattern of clonal expansion of pathogens followed by global spread of "super-fit" bacterial lineages (*32*). However, analysis of plasmids has shown that recombination of antibiotic resistance genes can occur, and entire plasmid sequences need to be studied to avoid drawing misleading conclusions (*33*). In addition, transmission of plasmids can occur within and even between genera, enhancing and promoting new epidemics (*34*, *35*). Our strategy relied on multiple methods to determine plasmid relationships and coupled it with ancestry of chromosomes to unravel the mosaicism of bacterial evolution. This approach has applications to other systems in modeling the impact of mobile genetic elements on bacterial evolution to limit risks from infectious diseases.

REFERENCES AND NOTES

1. B. A. McDonald, E. H. Stukenbrock, Rapid emergence of pathogens in agro-ecosystems: Global threats to agricultural sustainability and food security. *Philos. Trans. R. Soc. London Ser. B* **371**, 20160026 (2016). doi: 10.1098/rstb.2016.0026; pmid: 28080995
2. M. Carvajal-Yepes *et al*., A global surveillance system for crop diseases. *Science* **364**, 1237–1239 (2019). doi: 10.1126/science.aaw1572; pmid: 31249049
3. T. J. Johnson, L. K. Nolan, Pathogenomics of the virulence plasmids of *Escherichia coli*. *Microbiol. Mol. Biol. Rev.* **73**, 750–774 (2009). doi: 10.1128/MMBR.00015-09; pmid: 19946140
4. K. E. Holt *et al*., Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E3574–E3581 (2015). doi: 10.1073/pnas.1501049112; pmid: 26100894
5. S. Reuter *et al*., Parallel independent evolution of pathogenicity within the genus *Yersinia*. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 6768–6773 (2014). doi: 10.1073/pnas.1317161111; pmid: 24753568
6. D. Jamrozy *et al*., Evolution of mobile genetic element composition in an epidemic methicillin-resistant *Staphylococcus aureus*: Temporal changes correlated with frequent loss and gain events. *BMC Genomics* **18**, 684 (2017). doi: 10.1186/s12864-017-4065-z; pmid: 28870171
7. A. Orlek *et al*., Plasmid Classification in an Era of Whole-Genome Sequencing: Application in Studies of Antibiotic Resistance Epidemiology. *Front. Microbiol.* **8**, 182 (2017). doi: 10.3389/fmicb.2017.00182; pmid: 28232822
8. S. C. Slater *et al*., Genome sequences of three agrobacterium biovars help elucidate the evolution of multichromosome genomes in bacteria. *J. Bacteriol.* **191**, 2501–2511 (2009). doi: 10.1128/JB.01779-08; pmid: 19251847
9. E. W. Nester, *Agrobacterium*: Nature's genetic engineer. *Front. Plant Sci.* **5**, 730 (2015). pmid: 25610442
10. P. J. Christie, J. E. Gordon, The Agrobacterium Ti Plasmids. *Microbiol. Spectr.* **2**, PLAS-0010-2013 (2014). doi: 10.1128/microbiolspec.PLAS-0010-2013; pmid: 25610442
11. L.-Y. Lee, S. B. Gelvin, T-DNA binary vectors and systems. *Plant Physiol.* **146**, 325–332 (2008). doi: 10.1104/pp.107.113001; pmid: 18250230
12. M. F. Thomashow, S. Hugly, W. G. Buchholz, L. S. Thomashow, Molecular basis for the auxin-independent phenotype of crown gall tumor tissues. *Science* **231**, 616–618 (1986). doi: 10.1126/science.3511528; pmid: 3511528
13. Y. Dessaux, A. Petit, S. Farrand, P. Murphy, in *The Rhizobiaceae: Molecular Biology of Model Plant-Associated Bacteria* (Springer, 1998), pp. 173–197.
14. J. Tempe, P. Guyon, D. Tepfer, A. Petit, in *Plasmids of Medical, Environmental, and Commercial Importance*, K. H. Timmis, A. Pühler, Eds. (Elsevier/North Holland, 1979), pp. 353–363.
15. L. Otten *et al*., Evolution of agrobacteria and their Ti plasmids—A review. *Mol. Plant Microbe Interact.* **5**, 279–287 (1992). doi: 10.1094/MPMI-5-279; pmid: 1515664
16. K. Chen, L. Otten, Natural *Agrobacterium* Transformants: Recent Results and Some Theoretical Considerations. *Front. Plant. Sci.* **8**, 1600 (2017). doi: 10.3389/fpls.2017.01600; pmid: 28966626
17. J. M. Young, L. D. Kuykendall, E. Martínez-Romero, A. Kerr, H. Sawada, A revision of *Rhizobium* Frank 1889, with an emended description of the genus, and the inclusion of all species of *Agrobacterium* Conn 1942 and *Allorhizobium undicola* de Lajudie et al. 1998 as new combinations: *Rhizobium radiobacter, R. rhizogenes, R. rubi, R. undicola* and *R. vitis*. *Int. J. Syst. Evol. Microbiol.* **51**, 89–103 (2001). doi: 10.1099/00207713-51-1-89; pmid: 11211278
18. S. K. Farrand, P. B. van Berkum, P. Oger, *Agrobacterium* is a definable genus of the family *Rhizobiaceae*. *Int. J. Syst. Evol. Microbiol.* **53**, 1681–1687 (2003). doi: 10.1099/ijs.0.02445-0; pmid: 13130068
19. P. J. Keane, A. Kerr, P. B. New, Crown gall of stone fruit. II. Identification and nomenclature of *Agrobacterium* isolates. *Aust. J. Biol. Sci.* **23**, 585–595 (1970). doi: 10.1071/BI9700585
20. C. Mougel, J. Thioulouse, G. Perrière, X. Nesme, A mathematical method for determining genome divergence and species delineation using AFLP. *Int. J. Syst. Evol. Microbiol.* **52**, 573–586 (2002). doi: 10.1099/00207713-52-2-573; pmid: 11931171
21. H. Ochman, A. C. Wilson, Evolution in bacteria: Evidence for a universal substitution rate in cellular genomes. *J. Mol. Evol.* **26**, 74–86 (1987). doi: 10.1007/BF02111283; pmid: 3125340
22. R. Chriki-Adeeb, A. Chriki, Estimating Divergence Times and Substitution Rates in Rhizobia. *Evol. Bioinform. Online* **12**, EBO. S39070 (2016). doi: 10.4137/EBO.S39070; pmid: 27168719
23. V. L. Waters, K. H. Hirata, W. Pansegrau, E. Lanka, D. G. Guiney, Sequence identity in the nick regions of IncP plasmid transfer origins and T-DNA borders of *Agrobacterium* Ti plasmids. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 1456–1460 (1991). doi: 10.1073/pnas.88.4.1456; pmid: 1996345
24. E. G. Peralta, L. W. Ream, T-DNA border sequences required for crown gall tumorigenesis. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 5112–5116 (1985). doi: 10.1073/pnas.82.15.5112; pmid: 3860847
25. E. G. Peralta, R. Hellmiss, W. Ream, Overdrive, a T-DNA transmission enhancer on the *A. tumefaciens* tumour-inducing plasmid. *EMBO J.* **5**, 1137–1142 (1986). doi: 10.1002/j.1460-2075.1986.tb04338.x; pmid: 15966101
26. M. E. Wetzel, G. J. Olsen, V. Chakravartty, S. K. Farrand, The *repABC* plasmids with quorum-regulated transfer systems in members of the Rhizobiales divide into two structurally and separately evolving groups. *Genome Biol. Evol.* **7**, 3337–3357 (2015). doi: 10.1093/gbe/evv227; pmid: 26590210
27. M. C. John, R. M. Amasino, Expression of an *Agrobacterium* Ti plasmid gene involved in cytokinin biosynthesis is regulated by virulence loci and induced by plant phenolic compounds. *J. Bacteriol.* **170**, 790–795 (1988). doi: 10.1128/JB.170.2.790-795.1988; pmid: 2448293
28. D. E. Akiyoshi, H. Klee, R. M. Amasino, E. W. Nester, M. P. Gordon, T-DNA of *Agrobacterium tumefaciens* encodes an enzyme of cytokinin biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5994–5998 (1984). doi: 10.1073/pnas.81.19.5994; pmid: 6091129
29. N. Sardesai *et al*., Cytokinins secreted by *Agrobacterium* promote transformation by repressing a plant myb transcription factor. *Sci. Signal.* **6**, ra100 (2013). doi: 10.1126/scisignal.2004518; pmid: 24255177
30. H.-H. Hwang *et al*., The Tzs protein and exogenous cytokinin affect virulence gene expression and bacterial growth of *Agrobacterium tumefaciens*. *Phytopathology* **103**, 888–899 (2013). doi: 10.1094/PHYTO-01-13-0020-R; pmid: 23593941
31. E. A. Savory *et al*., Evolutionary transitions between beneficial and phytopathogenic *Rhodococcus* challenge disease

management. *eLife* **6**, e30925 (2017). doi: 10.7554/eLife.30925; pmid: 29231813
32. S. Baker, N. Thomson, F.-X. Weill, K. E. Holt, Genomic insights into the emergence and spread of antimicrobial-resistant bacterial pathogens. *Science* **360**, 733–738 (2018). doi: 10.1126/science.aar3777; pmid: 29773743
33. A. E. Sheppard *et al*., Nested Russian Doll-Like Genetic Mobility Drives Rapid Dissemination of the Carbapenem Resistance Gene bla_{KPC}. *Antimicrob. Agents Chemother.* **60**, 3767–3778 (2016). doi: 10.1128/AAC.00464-16; pmid: 27067320
34. S. Conlan *et al*., Single-molecule sequencing to track plasmid diversity of hospital-associated carbapenemase-producing Enterobacteriaceae. *Sci. Transl. Med.* **6**, 254ra126 (2014). doi: 10.1126/scitranslmed.3009845; pmid: 25232178
35. K. S. Baker *et al*., Horizontal antimicrobial resistance transfer drives epidemics of multiple *Shigella* species. *Nat. Commun.* **9**, 1462 (2018). doi: 10.1038/s41467-018-03949-8; pmid: 29654279

ACKNOWLEDGMENTS

We thank members of the community for their support, their eagerness in sharing ideas, data, and resources, as well as their lively discussions at Crown Gall Conferences. We thank L. Moore for foresight and endeavors in preserving the agrobacterial culture collection, and the many growers, S. Farrand, and J. Puławska for providing samples and strains. We also thank J. Anderson, T. Sharpton, C. Meehan, E.-M. Lai, W. Ream, and members of the Chang, Putnam (Plant Clinic), Loper, and Grünwald labs for their assistance and insightful comments. We acknowledge the staff of the CGRB for their services. **Funding:** Supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture, award 2014-51181-22384 (J.H.C., M.L.P., and N.J.G.); intramural funding from Academia Sinica (C.-H.K.); USDA NIFA award 2017-67012-26126 (A.J.W.); and a Provost's Distinguished Graduate Fellowship awarded by Oregon State University (E.W.D.). This material is based on work supported by a NSF Graduate Research Fellowship under grant DGE-1314109 to E.W.D. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. **Author contributions:** A.J.W., E.W.D., J.T., M.M., J.E.L., N.J.G., M.L.P., and J.H.C. conceptualized and designed the experiments. A.J.W., E.W.D., J.T., and M.S.B. performed the research. A.J.W., E.W.D., J.T., N.J.G., M.L.P., and J.H.C. analyzed the data. C.-H.K. contributed resources prior to their publication. A.J.W., E.W.D., J.T., C.-H.K., N.J.G., M.L.P., and J.H.C. wrote the paper. A.J.W., E.W.D., C.-H.K., N.J.G., M.L.P., and J.H.C. acquired funding. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Short reads and assemblies have been deposited in NCBI as BioProject PRJNA607555 and accession numbers are listed in data S1 and S10. Network graphs in nexus or sif format, phylogenetic trees in Newick format, genome annotations, and scripts can be downloaded from https://github.com/osuchanglab/AgrobacteriumEvolutionManuscript (DOI:10.5281/zenodo.3754985). Strains sequenced in this study are available upon request. Records were deidentified to mitigate privacy risks to growers.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6495/eaba5256/suppl/DC1
Materials and Methods
Figs. S1 to S25
Data S1 to S19
References (*36*–*86*)

11 December 2019; resubmitted 28 February 2020
Accepted 27 April 2020
10.1126/science.aba5256

RESEARCH ARTICLE

STRUCTURAL BIOLOGY

The structure of human CST reveals a decameric assembly bound to telomeric DNA

Ci Ji Lim[1,2], Alexandra T. Barbour[1], Arthur J. Zaug[1,2,3], Karen J. Goodrich[1,2,3], Allison E. McKay[1,2], Deborah S. Wuttke[1]*, Thomas R. Cech[1,2,3]*

The CTC1-STN1-TEN1 (CST) complex is essential for telomere maintenance and resolution of stalled replication forks genome-wide. Here, we report the 3.0-angstrom cryo–electron microscopy structure of human CST bound to telomeric single-stranded DNA (ssDNA), which assembles as a decameric supercomplex. The atomic model of the 134-kilodalton CTC1 subunit, built almost entirely de novo, reveals the overall architecture of CST and the DNA-binding anchor site. The carboxyl-terminal domain of STN1 interacts with CTC1 at two separate docking sites, allowing allosteric mediation of CST decamer assembly. Furthermore, ssDNA appears to staple two monomers to nucleate decamer assembly. CTC1 has stronger structural similarity to Replication Protein A than the expected similarity to yeast Cdc13. The decameric structure suggests that CST can organize ssDNA analogously to the nucleosome's organization of double-stranded DNA.

CTC1-STN1-TEN1 (CST) is a protein complex essential for telomere replication (*1*–*4*) and as a DNA polymerase alpha-primase (Pol-α) cofactor (*5*), and it functions genome-wide to recover stalled replication forks (*2*, *6*–*9*) and facilitate DNA damage repair (*10*–*13*). Consequently, mutations in CST are the basis of human genetic diseases such as Coats plus syndrome and dyskeratosis congenita (*14*–*19*).

Although it preferentially binds short telomeric single-stranded DNA (ssDNA) (*20*–*23*), CST can also bind less specifically to longer ssDNA (*2*, *4*). An intact heterotrimeric CST complex is necessary for its DNA-binding function (*4*, *15*, *24*), but limited understanding of mammalian CST architecture has hampered the determination of its DNA-binding region(s). Structures of human components are limited to STN1 and TEN1 (*24*), and solving the structure of the largest subunit CTC1 has been technically challenging, with only a single domain being determined (*25*). The yeast Cdc13 protein associates with Stn1 and Ten1 and has therefore been proposed as a CTC1 homolog, despite Cdc13 and mammalian CTC1 being unrelated in sequence. Hence, it has been unclear if Cdc13 and CTC1 share structural homology.

Cryo–electron microscopy (cryo-EM) structure of human CST decameric supercomplex

We solved the structure of purified recombinant human CST protein (hereafter termed "DNA-free CST") to 6.3-Å resolution using single-particle cryo-EM (fig. S1). To improve the resolution of the structure, we added a minimal telomeric ssDNA [3xTEL, $(TTAGGG)_3$] (*4*) to the purified CST protein and unexpectedly discovered a symmetric complex that was considerably larger than the monomeric CST (Fig. 1A and fig. S2, A and B). Subsequent cryo-EM processing revealed that the symmetric complex was a decameric supercomplex (10 CST monomers) with D5 symmetry, which was reconstructed at 3.0-Å global resolution (Fig. 1A and fig. S2C). The CST monomers were computationally extracted from the supercomplex and further sorted to obtain a final set of data, which led to the cryo-EM map of a CST monomer still at 3.0-Å global resolution (Fig. 1B and figs. S2D and S3 to S5) but with the map quality substantially improved (fig. S5A). This enabled us to dock all the available crystal structures of the domains of human TEN1, STN1 (*24*), and a central OB (oligonucleotide-oligosaccharide–binding fold) domain of CTC1 (*25*) with high confidence (Fig. 1B). Furthermore, we were able to build de novo the remaining unsolved body of CTC1 [894 residues, excluding the one CTC1 OB domain previously solved (*25*)] (fig. S5B).

Human chromosome ssDNA telomeric overhangs are 50 to 200 nucleotides (nt) long, much larger than 3xTEL, so we tested CST binding to 15xTEL (90 nt). CST bound 15xTEL with sixfold higher affinity than 3xTEL (fig. S6, A to C, and table S1), and decameric CST supercomplexes were readily apparent by negative-stain EM (fig. S6D). Thus, the decamer can form with both long and short telomeric ssDNA molecules.

Overall architecture of human CST

Model building revealed the overall architecture of the human CST heterotrimer (Fig. 1, C and D; fig. S5B; and table S2). CTC1 is composed of seven tandem OB domains (OB-A through G; Fig. 1, C and D). The human CST complex has a subunit stoichiometry of 1:1:1, unlike the nonuniform stoichiometry reported for the *Candida glabrata* CST complex (*26*).

The C terminus of CTC1 (OB-D through G) serves as a hub for STN1 and TEN1 assembly (Fig. 1D). A single STN1 protein has two separate interaction sites with CTC1, with the STN1 N-terminal half (STN1n) interacting with CTC1 OB-G and the C-terminal half (STN1c) with CTC1 OB-E (Fig. 2A). These two halves of STN1 are connected by an unstructured peptide linker of seven residues (Fig. 2A). In contrast to the related ssDNA-binding protein, Replication Protein A (RPA) (*27*, *28*), there is no triple-helix bundle stabilizing the heterotrimeric CST complex (Fig. 2B). Instead, TEN1 binding to CTC1 is bridged by STN1n [similar to the model of the *Tetrahymena* CST (*29*, *30*)] (Fig. 2B), with STN1n binding to a highly conserved interaction patch on CTC1 OB-G (fig. S7A).

The first winged helix-turn-helix (wHTH) domain of STN1c interacts with CTC1 OB-E (Fig. 1D). However, no strong conservation of residues occurs on the interaction patch of CTC1 OB-E (fig. S7A), suggesting that STN1c-CTC1 interaction could be weaker than STN1n-CTC1 interaction, as reported for the *Tetrahymena* CST complex (*30*). Supporting this hypothesis, we found that STN1n alone was able to interact with CTC1, but STN1c could not (fig. S7, B and C). In addition, TEN1 interaction with CTC1 was maintained with STN1n but lost when only STN1c was present. STN1n and CTC1 interact through two regions—CTC1 "cleft region" (the conserved patch on CTC1; Fig. 2B and fig. S7A) and a new CTC1-STN1n three-helix bundle (Fig. 2C). The importance of the cleft and the three-helix bundle for CTC1-STN1 association was confirmed by mutagenesis (fig. S7, D and E).

CTC1 OB folds E, F, and G are arranged spatially on OB-D, which acts like a scaffold, resulting in these four OBs forming a ringlike structure (Fig. 1D). Structural homology analysis of individual CTC1 OB domains found CTC1 to be most similar to RPA and Teb1 (an RPA-like paralog in *Tetrahymena*) (fig. S8), with CTC1 OB-F most similar to Teb1's OB-B (*31*) and OB-G similar to OB-C of RPA70 or Teb1 (*27*). CTC1 OB-G also has a conserved zinc ribbon motif like that of the OB-C domains of RPA and Teb1 (*27*, *31*) (fig. S9). The scaffolding OB-D has no convincing structural homologies, but given its distinctive extended OB-fold structure (Fig. 1D), it could be an evolved form of the more compact and conventional OB-fold (*32*). Notably, despite the long-standing suggestion that yeast Cdc13 and mammalian CTC1 are homologs (*33*), we found weaker structural homologies to Cdc13 than with the best RPA70 homology matches (based on DALI structural homology Z-score, fig. S10).

[1]Department of Biochemistry, University of Colorado Boulder, Boulder, CO 80303, USA. [2]BioFrontiers Institute, University of Colorado Boulder, Boulder, CO 80303, USA. [3]Howard Hughes Medical Institute, University of Colorado Boulder, Boulder, CO 80303, USA.
*Corresponding author. Email: thomas.cech@colorado.edu (T.R.C.); deborah.wuttke@colorado.edu (D.S.W.)

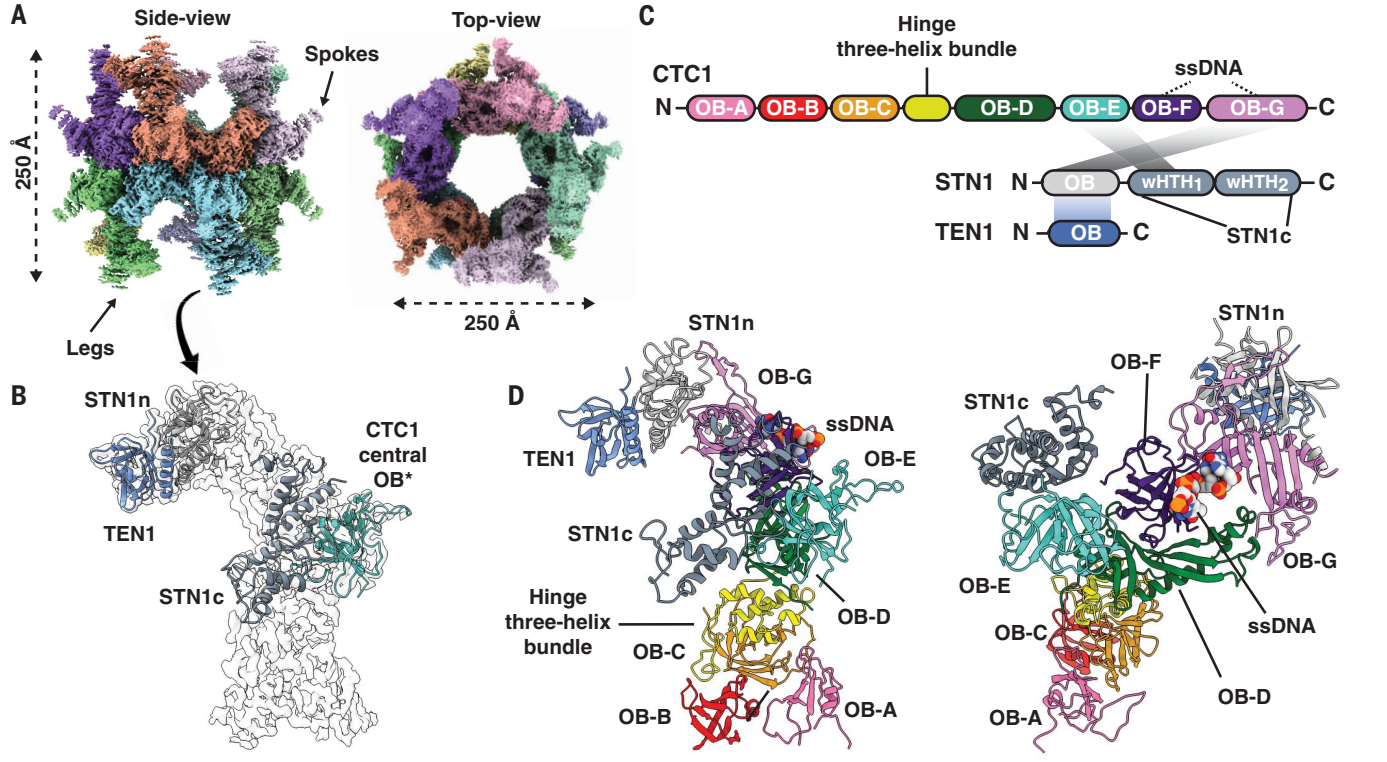


Fig. 1. Cryo-EM structure of human CST decameric supercomplex and its architecture. (**A**) Cryo-EM density of decameric CST complex colored by segmented CST monomers. (**B**) Docking of available atomic models of a CTC1 OB-domain [*reported as central domain OB-fold (*25*), PDB 5W2L], STN1n (N-terminal half, PDB 4JOI:A), STN1c (C-terminal half, PDB 4JQF), and TEN1 (PDB 4JOI:C). (**C**) Structure-based schematic of CST domain architecture and intermolecular interactions between subunits. The individual OB domains of CTC1 are rainbow colored. (**D**) CTC1 architectural organization of seven OB domains (A to G) and the identified bound-ssDNA (space-filled model). Spokes in (A) are STN1c, whereas the legs are CTC1 OB-A, -B, and -C.

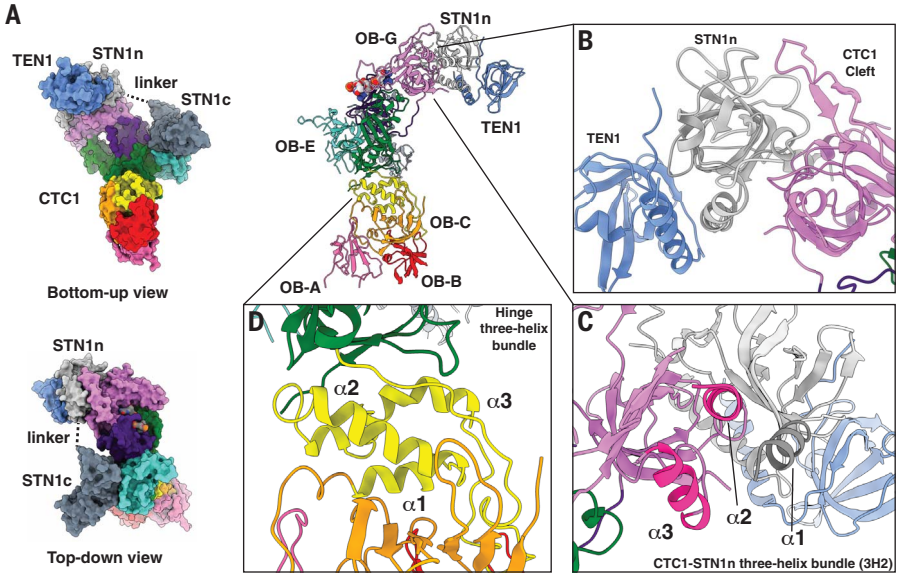


Fig. 2. CST intersubunit interactions and CTC1 molecular motifs. (**A**) STN1 N-terminal and C-terminal halves—STN1n and STN1c—interact separately with CTC1 by means of a flexible peptide linker. (**B**) STN1 and TEN1 do not interact with CTC1 using a trimeric helix-bundle like human RPA; instead, STN1 directly interacts with a highly conserved patch on CTC1 (Cleft) and bridges TEN1 to CTC1. (**C**) CTC1-STN1n three-helix bundle that is involved in CTC1 and STN1 assembly. α1 (highlighted dark gray) is from STN1n and α2 and α3 (highlighted bright pink) are from CTC1 OB-G. (**D**) The hinge three-helix bundle (annotated α1, α2, and α3) connects OB-A, -B, and -C to the rest of the C-terminal OB-domains of CTC1.

We also found an intramolecular three-helix bundle bridging OB-C and OB-D (Fig. 2D). This three-helix bundle (termed hinge three-helix bundle) effectively segregates OB-A, -B, and -C from the C-terminal OB domains (Fig. 1, C and D). CST OB-C serves as a scaffold for both OB-A and OB-B (Fig. 1D). Because of extensive flexibility of OB-A, we could only de novo build a poly(alanine) model (~45 residues) for it, with the overall backbone of OB-A clearly showing the OB-fold topology. Structural homology searches of OB-B and OB-C reveal them to be most similar to *Ustilago maydis* RPA70 OB-A and OB-B (*28*) (fig. S8). The multiple structural homologies of human CTC1 to various domains of RPA70 suggest that CTC1 may have evolved from RPA (figs. S8 and S10).

Disease mutations in CTC1 that have been shown to interfere with Pol-α binding (*15*) are located on CTC1 OB-B (A227 and V259) and on scaffold OB-D (V665) (fig. S11). Given that Pol-α has a bilobal architecture (*34*), the catalytic and primase lobes of Pol-α could engage CST at separate sites.

CST ssDNA-binding anchor site

Four nucleotides, TAGG, were clearly visible in the cryo-EM map of the complex and not in the DNA-free CST cryo-EM map (Fig. 3, A to

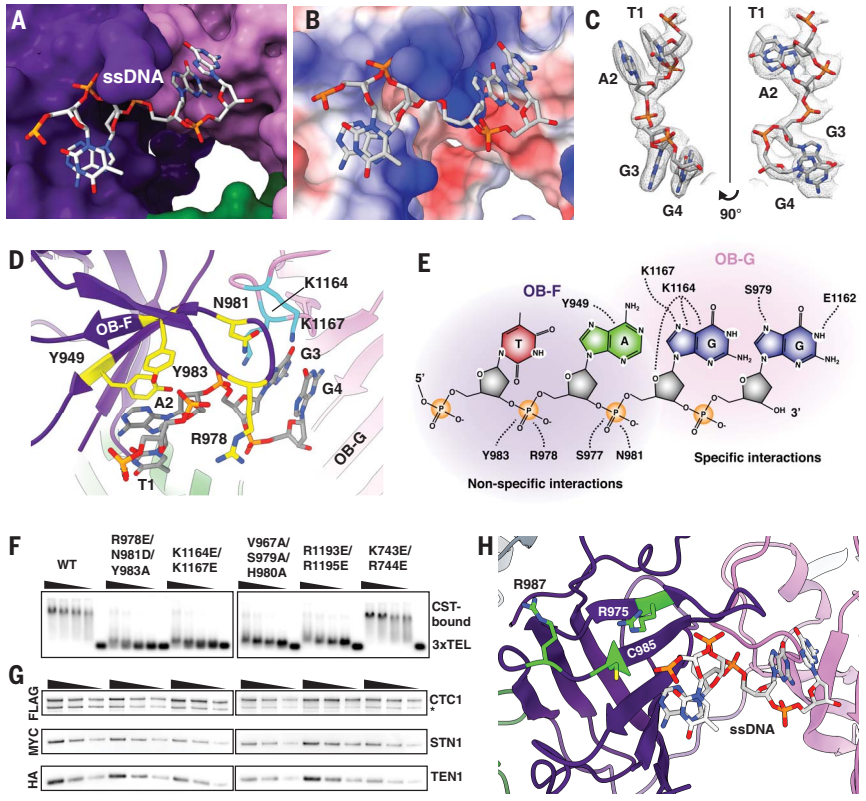


Fig. 3. Telomeric ssDNA-binding anchor site of CST. (**A**) A 4-nt segment of the single-stranded telomeric DNA is located on CTC1. (**B**) Coulombic surface analysis (*36*) reveals that the ssDNA anchor site is highly positively charged (blue; red is negatively charged surface). (**C**) Cryo-EM density of the ssDNA molecule built with the sequence assigned as TAGG (5′-T1-A2-G3-G4-3′). The numbering is based on the visible ssDNA, not the full-length ssDNA. (**D**) CTC1 residues involved in ssDNA binding are shown in yellow and cyan from OB-F (yellow on purple) and OB-G (cyan on pink), respectively. (**E**) Schematic of CST ssDNA-binding anchor site across CTC1 OB-F and OB-G. (**F**) Gel-shift assay showing that CST DNA-binding mutants predicted from the atomic model no longer bind telomeric ssDNA $(TTAGGG)_3$. Wedges indicate twofold dilutions of CST starting at 50 nM, with the fifth lane of each group having no protein added. K743E/R744E mutant does not directly bind DNA and was used as a control to test if charge swaps in the vicinity might be sufficient to destabilize DNA binding. (**G**) CST DNA-binding mutants can still form heterotrimeric CST complex as shown by tandem immunoprecipitation pull-down assays [FLAG/hemagglutinin (HA)] from exogenously expressed FLAG-CTC1, MYC-STN1, and HA-TEN1. Asterisk (*) indicates protein degradation product. Wedges indicate a twofold dilution that is used to ensure that Western-blot band intensities are in the linear detection range. (**H**) Human CTC1 disease mutations (*15*) that abolish ssDNA binding (lime green residues) are located near the ssDNA anchor site. Abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; G, Gly; H, His; K, Lys; L, Leu; N, Asn; R, Arg; S, Ser; V, Val; W, Trp; and Y, Tyr.

C; see table S3 for cross-correlation analysis, fig. S12A), suggesting that these interact with the protein most strongly or with highest occupancy. The rest of the ssDNA is likely flexible or bound to CST in multiple binding modes, which is consistent with CST being able to bind multiple configurations of ssDNA dynamically (*4*, *7*, *23*). Hereafter, we identify the site of ssDNA binding on CTC1 as the ssDNA anchor patch.

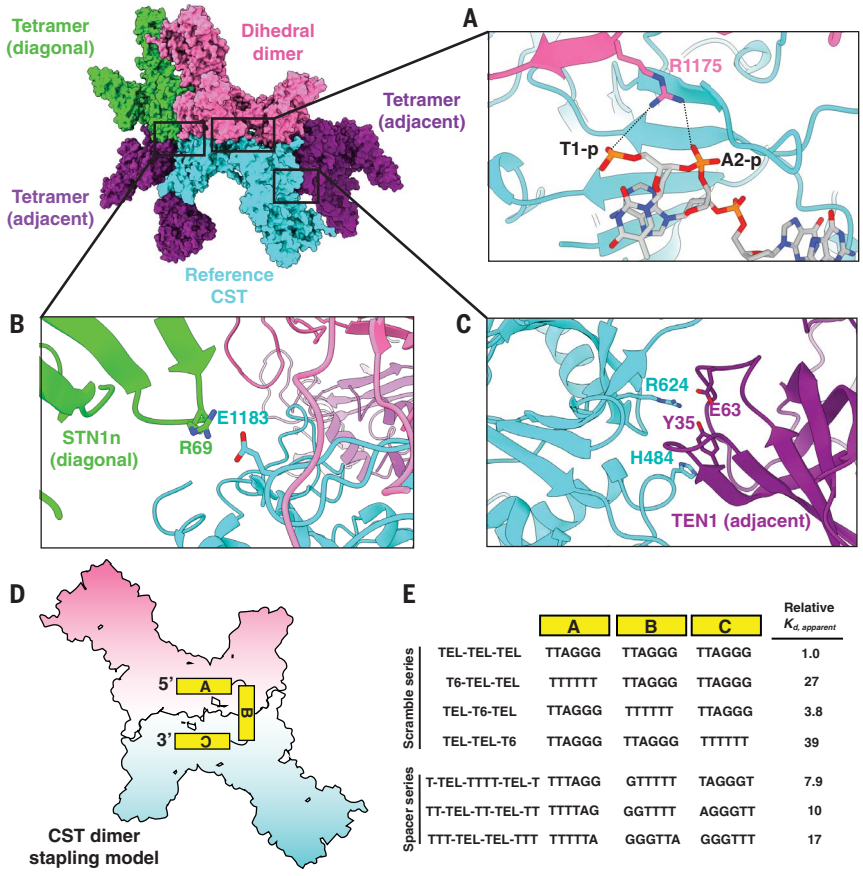
Several positively charged residues of CTC1 are involved in the interaction with ssDNA at the anchor patch (CTC1 R978, K1164, and K1167), as well as additional aromatic and neutral-polar residues (CTC1 Y949, N981, and Y983) (Fig. 3D and E). This anchor patch uses several kinds of interaction between CTC1 and ssDNA: for example, R978 and N981 hydrogen bonding to the negatively charged ssDNA phosphate backbone (fig. S12B); K1164 and K1167 hydrogen bonding to ssDNA bases (fig. S12C); and Y949 π-π stacking with the A2 base, which in turn is stacked on the T1 base (fig. S12D). The tyrosine-base-base stack is reminiscent of the stacking arrangements seen in several OB fold–nucleic acid interactions, e.g., the human POT1-ssDNA structure (*20*). The nonspecific interactions mostly involve CTC1 OB-F, whereas specific interactions are in OB-G (Fig. 3E); the modeled 4-nt ssDNA spans these two OB domains, suggesting that ssDNA binding stabilizes CTC1 architecture (Fig. 3D). In addition, these ssDNA-interacting residues are highly conserved across mammalian CTC1 homologs (fig. S13).

To validate the observed protein-DNA interactions, we performed mutagenesis on sets of CTC1 residues in the ssDNA anchor patch—R987E/N981D/Y983A (anchor site on OB-F), K1164E/K1167E (anchor site on OB-G), V967A/S979A/H980A (structural integrity residues on OB-F), and R1193E/R1195E (structural integrity residues on OB-G) (see fig. S14 for structural mapping of additional tested mutants). Each of these sets of mutations abolished CST DNA-binding activity, whereas the K743E/R744E negative control mutation did not (Fig. 3F). In all of these mutants, CST still forms a heterotrimer complex (Fig. 3G). Previously identified CTC1 disease mutations (R975G, C985Δ and R987W) that have been shown (*15*) to affect CST-ssDNA binding are also in the vicinity of the ssDNA anchor patch (Fig. 3H).

Assembly mechanism and pathways of decameric CST supercomplex

We identified interactions that appear to mediate decameric supercomplex assembly. The sites can be categorized into two oligomerization classes (Fig. 4, A to C): (i) dimerization (dihedral dimerization) and (ii) tetramerization (two subclasses: adjacent and diagonal). For CST dimerization, three conserved residues at the interface are N745, L843, and R1175. R1175 is particularly interesting, given that it is also within range (<5 Å) for interaction with the phosphodiester groups of T1 or A2 of the opposite dihedral dimer's telomeric ssDNA (Fig. 4A), which suggests ssDNA binding can also stabilize CST dihedral dimerization. Consistent with this prediction, CST with the CTC1 R1175E mutation showed a 26-fold reduction in DNA-binding ability with 3xTEL ssDNA but no effect when tested with a non-specific T18 (poly-T) ssDNA (fig. S15).

For tetramerization, CTC1 interacts with its diagonally opposite neighbor's STN1n (CTC1 E1183) (Fig. 4B) and adjacent neighbor's TEN1 (CTC1 H484 and R624) (Fig. 4C). The proximity of the two ssDNA anchor patches across dihedral dimers (fig. S16A) suggested that a single ssDNA molecule of three TTAGGG repeats could "staple" together two monomers into a dihedral dimer, with the first and last repeat engaged by the CST monomers while the middle repeat served as a linker (Fig. 4D). Consistent with this model, we found that replacing individual TTAGGG modules with T_6 reduced CST DNA-binding affinity for either the first or last repeat but was tolerated for the middle repeat (Fig. 4E and fig. S16, B and C). As an additional test, shortening the middle repeat sequence (using oligo-T sequence instead of TTAGGG) to <6 nt negatively affected CST-DNA binding (Fig. 4E and fig. S16, D and E), consistent with the measured molecular distance (~20 Å, fig. S16A) between the



		A	B	C	Relative $K_{d,\,apparent}$
Scramble series	TEL-TEL-TEL	TTAGGG	TTAGGG	TTAGGG	1.0
	T6-TEL-TEL	TTTTTT	TTAGGG	TTAGGG	27
	TEL-T6-TEL	TTAGGG	TTTTTT	TTAGGG	3.8
	TEL-TEL-T6	TTAGGG	TTAGGG	TTTTTT	39
Spacer series	T-TEL-TTTT-TEL-T	TTTAGG	GTTTTT	TAGGGT	7.9
	TT-TEL-TT-TEL-TT	TTTTAG	GGTTTT	AGGGTT	10
	TTT-TEL-TEL-TTT	TTTTTA	GGGTTA	GGGTTT	17

Fig. 4. Molecular interactions underlying CST decameric supercomplex formation and testing the dimer stapling model. (**A** to **C**) The reference CST (cyan) is flanked by four CST complexes—a dihedral dimer (opposite, pink) and three tetrameric partners (one diagonal neighbor, green, and two adjacent neighbors, purple). (A) CTC1 R1175 from the dihedral dimer neighbor (pink) is pointing toward ssDNA bound to the reference CTC1 (cyan), with the black dashed lines representing feasible ionic interactions between R1175 and phosphodiester groups of the ssDNA. (B and C) Identified intermolecular interactions between CTC1, STN1, and TEN1 at interfaces of the decameric supercomplex. (**D**) CST dimer stapling model with an 18-nt ssDNA molecule. The two monomers are separately colored as pink and cyan for visual clarity. (**E**) Changes in CST DNA-binding affinity ($K_{d,apparent}$) relative to that of 3xTEL with oligo-T substitution of block A, B, or C of 3xTEL (Scramble series). The molecular distance between the TTAGGG sequences of blocks A and C was also varied, and the impact on CST relative DNA-binding affinity was measured (Spacer series). TEL-TEL-TEL oligo is also known as 3xTEL. The relative DNA-binding affinity values are reported to two significant figures; measured values and error analysis are in table S1.

5′ and 3′ ends of the neighboring ssDNA molecules in the dihedral CST dimer.

A comparison of the DNA-free CST model and the monomeric CST model extracted from the decameric supercomplex (for example, compare fig. S1D to fig. S2D) suggested that STN1c has two alternate docking sites on CTC1. To investigate this, we turned to a cryo-EM dataset that had a high population of monomeric CST with telomeric ssDNA added (fig. S17) and found two distinct conformations—one with a "head" density and the other with an "arm" density—albeit at a lower model resolution of ~9 Å (Fig. 5, A and B). Because STN1c is in the "arm" conformation in the decameric CST, and the STN1c "head" conformation would sterically hinder formation of the decamer (by obstructing dihedral dimerization), we propose that switching from "head" to "arm" docking position for STN1c is an important first step for CST to form a decameric supercomplex. STN1c switching is consistent with our finding that STN1c is less stably bound to CTC1 than STN1n (fig. S7, B and C).

Surface-charge analysis revealed a highly positively charged surface on CTC1 OB-G, where the STN1c is expected to dock in the "head" conformation (Fig. 5C), and similar analysis revealed a reciprocal patch of high negative charge on STN1c (Fig. 5C, inset). This suggested that charge-charge interactions could mediate the transition from monomeric to decameric CST, explaining how a longer ssDNA, with extended binding to the OB-G's negative patch, can trigger this transition. The charge-charge interactions also suggested that increased salt concentration could mediate the transition in the absence of ssDNA. Indeed, we found a large increase in decameric CST population without addition of ssDNA in a nonphysiological salt concentration of 800 mM NaCl (fig. S18).

Finally, we used negative-stain EM single-particle analysis to identify subcomplexes of the decamer that would give hints to its assembly pathway(s). We observed two subcomplexes, dimers and tetramers (Fig. 5D), which are plausible intermediates in an assembly pathway based on ssDNA-stapled dimers such as the following: CST assembles first as a dihedral dimer before forming a lateral tetramer involving two dihedral dimers, and sequential addition of dimers eventually closes the symmetric circle (decamer) by continuing the lateral oligomerization (Fig. 5E).

Evidence for higher-order CST assemblies in vivo

CST monomers interact specifically to form the decamer, burying a great deal of exposed protein surface area (~2100 Å^2 per monomer), and ssDNA has a specific role in triggering decamer assembly. These features indicate that formation of the CST decamer is thermodynamically favorable and that the monomer is built to form the decamer. To confirm that CST forms oligomers in cells, we turned to an orthogonal epitope tag pull-down approach. V5-tagged and FLAG-tagged CTC1 were coexpressed in human embryonic kidney 293T (HEK293T) cells, along with STN1 and TEN1. Pull-down using anti-FLAG beads immunoprecipitated V5-CTC1, as well as FLAG-CTC1, and the reciprocal experiment with anti-V5 beads similarly recovered the CTC1 with both epitope tags (Fig. 5, F and G, and fig. S19, A and B, for IP controls). Notably, the pull-down result was not sensitive to DNA and RNA degradation with benzonase (Fig. 5, F and G, and fig. S19C), so the higher-order complexes were not loosely tethered by nucleic acid. We conclude that a substantial fraction of CST resides in a higher-order protein complex, consistent with decamers existing in vivo.

Discussion

Our structure of the decameric human CST supercomplex bound with telomeric ssDNA provides the platform for understanding mechanisms of various CST functions in DNA replication and DNA damage repair, not only at telomeres but also genome-wide (*7*, *8*, *11*, *13*, *35*). The atomic-resolution details revealed in this structure enabled us to identify amino acids

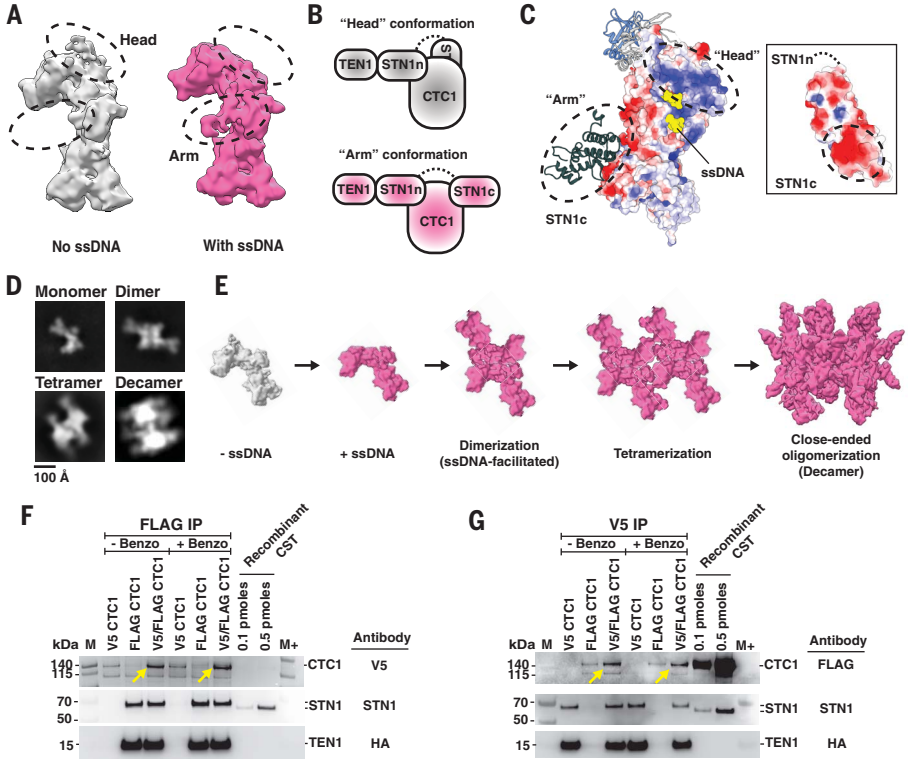


Fig. 5. Assembly mechanism and pathway model of CST decameric supercomplex. (**A**) Cryo-EM densities of two conformations of monomeric CST with the differences indicated by dashed black circles. The two conformations—"head" (colored gray) and "arm" (colored pink)—are assigned as CST without and with ssDNA bound, respectively. (**B**) Cartoon models of CST "head" and "arm" conformations depicted by conformational changes of STN1c docking site on CTC1. The black dashed line represents the unstructured polypeptide region between STN1n and STN1c. (**C**) Coulombic surface analysis reveals a highly positively charged patch on CTC1 OB-G, where STN1c lies when in "head" conformation. Reciprocally, a highly negatively charged surface is shown on STN1c (see inset). (**D**) Two-dimensional class averages of negative-stained CST incubated with 3xTEL ssDNA showed multiple oligomeric species of CST, which are assigned as monomer, dimer, tetramer, and decamer. (**E**) Proposed model of assembly pathway of CST decameric supercomplex upon ssDNA introduction. CST binding of ssDNA prevents STN1c from binding to its original site ("head" conformation, gray), allowing CST to form dimers before progressing to tetramers, and eventually leading to a close-ended decameric supercomplex. (**F** and **G**) Immunoprecipitation (IP) of orthogonally tagged CTC1 molecules coexpressed in cells. (F) FLAG IP of HEK293T cell extracts that were cotransfected with V5-CTC1, FLAG-CTC1, or both, and with TEN1 and STN1. Western blot with antibody against V5 showed that FLAG-IP of FLAG-CTC1 also coimmunoprecipitated V5-CTC1 (yellow arrows). (G) Coimmunoprecipitation of FLAG-CTC1 was also observed with V5-IP of V5-CTC1 (yellow arrows). STN1 and TEN Western blots were done to determine the presence of CST heterotrimeric complex assembly. M and M+ indicate protein ladder PageRuler and PageRuler Plus, respectively.

responsible for the numerous interactions that are key for CST functions, i.e., the intricate heterotrimer assembly, ssDNA-binding anchor site, and decamer assembly. Moreover, the structure provides a molecular model to understand the underlying mechanisms of CST mutants in human diseases such as Coats plus and dyskeratosis congenita. Finally, we speculate that the decamer could be a nucleosome-equivalent for G-rich ssDNA, compacting it and competing with G-quadruplex structures both at stalled replication forks and at telomeres.

REFERENCES AND NOTES

1. Y. V. Surovtseva *et al.*, *Mol. Cell* **36**, 207–218 (2009).
2. Y. Miyake *et al.*, *Mol. Cell* **36**, 193–206 (2009).
3. F. Wang *et al.*, *Cell Rep.* **2**, 1096–1103 (2012).
4. L. Y. Chen, S. Redon, J. Lingner, *Nature* **488**, 540–544 (2012).
5. D. E. Casteel *et al.*, *J. Biol. Chem.* **284**, 5807–5818 (2009).
6. M. Zhang *et al.*, *Nucleic Acids Res.* **47**, 5243–5259 (2019).
7. A. Bhattacharjee, Y. Wang, J. Diao, C. M. Price, *Nucleic Acids Res.* **45**, 12311–12324 (2017).
8. M. Chastain *et al.*, *Cell Rep.* **16**, 1300–1314 (2016).
9. J. A. Stewart *et al.*, *EMBO J.* **31**, 3537–3549 (2012).
10. K. Kratz, T. de Lange, *J. Biol. Chem.* **293**, 14384–14392 (2018).
11. Z. Mirman *et al.*, *Nature* **560**, 112–116 (2018).
12. M. Greetham, E. Skordalakes, D. Lydall, B. A. Connolly, *J. Mol. Biol.* **427**, 3023–3030 (2015).
13. F. Wang, J. Stewart, C. M. Price, *Cell Cycle* **13**, 3488–3498 (2014).
14. J. A. Stewart, Y. Wang, S. M. Ackerson, P. L. Schuck, *Front. Biosci.* **23**, 1564–1586 (2018).
15. L. Y. Chen, J. Majerská, J. Lingner, *Genes Dev.* **27**, 2099–2108 (2013).
16. M. Armanios, *Pediatr. Blood Cancer* **59**, 209–210 (2012).
17. B. H. Anderson *et al.*, *Nat. Genet.* **44**, 338–342 (2012).
18. M. Amir *et al.*, *Front. Mol. Biosci.* **6**, 41 (2019).
19. P. Gu, S. Chang, *Aging Cell* **12**, 1100–1109 (2013).
20. M. Lei, E. R. Podell, T. R. Cech, *Nat. Struct. Mol. Biol.* **11**, 1223–1229 (2004).
21. D. Loayza, H. Parsons, J. Donigian, K. Hoke, T. de Lange, *J. Biol. Chem.* **279**, 13241–13248 (2004).
22. P. Baumann, T. R. Cech, *Science* **292**, 1171–1175 (2001).
23. R. A. Hom, D. S. Wuttke, *Biochemistry* **56**, 4210–4218 (2017).
24. C. Bryan, C. Rice, M. Harkisheimer, D. C. Schultz, E. Skordalakes, *PLOS ONE* **8**, e66756 (2013).
25. P. K. Shastrula, C. T. Rice, Z. Wang, P. M. Lieberman, E. Skordalakes, *Nucleic Acids Res.* **46**, 972–984 (2018).
26. N. F. Lue *et al.*, *PLOS Genet.* **9**, e1003145 (2013).
27. E. Bochkareva, S. Korolev, S. P. Lees-Miller, A. Bochkarev, *EMBO J.* **21**, 1855–1863 (2002).
28. J. Fan, N. P. Pavletich, *Genes Dev.* **26**, 2337–2347 (2012).
29. B. Wan *et al.*, *Nat. Struct. Mol. Biol.* **22**, 1023–1026 (2015).
30. J. Jiang *et al.*, *Science* **350**, aab4070 (2015).
31. Z. Zeng *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20357–20361 (2011).
32. D. L. Theobald, R. M. Mitton-Fry, D. S. Wuttke, *Annu. Rev. Biophys. Biomol. Struct.* **32**, 115–133 (2003).
33. C. Rice, E. Skordalakes, *Comput. Struct. Biotechnol. J.* **14**, 161–167 (2016).
34. R. Núñez-Ramírez *et al.*, *Nucleic Acids Res.* **39**, 8187–8199 (2011).
35. L. Y. Chen, J. Lingner, *Nucleus* **4**, 277–282 (2013).
36. E. F. Pettersen *et al.*, *J. Comput. Chem.* **25**, 1605–1612 (2004).

ACKNOWLEDGMENTS

We thank Z. Yu, D. Matthies and R. Huang (Janelia Research Campus), P. Blerkom and J. Kieft (University of Colorado Anschutz), and G. Morgan and C. Page (University of Colorado Boulder) for microscope setup and data collection. We thank F. Asturias (University of Colorado Anschutz) and T. Terwilliger (Los Alamos National Laboratory) for discussion. We thank T. Nahreini (University of Colorado Boulder) and BioFrontiers Institute Computing Core for their support and assistance. We thank D. Youmans for input regarding immunoprecipitation experiments. In addition, we thank the members of the Cech and Wuttke laboratories for their suggestions. **Funding:** This work was funded in part by grants from NIH to T.R.C. (R01GM099705), D.S.W. (R01GM059414), and C.J.L. (K99GM131023) and from NSF to D.S.W. (MCB 1716425). A.T.B. is supported by a fellowship provided by NIH–University of Colorado Boulder (T32GM008759). T.R.C. is an HHMI Investigator. **Author contributions:** C.J.L., D.S.W., and T.R.C. outlined the experimental plans. C.J.L. established experimental procedures, expressed and purified the complexes from insect cells, optimized conditions for cryo-EM work, prepared cryo-EM grids, collected and processed EM datasets, and did the model building, refinement and validations. A.T.B. assisted in insect cell expression and purification and performed negative-stain screening and analysis. A.J.Z. and A.E.M. performed molecular cloning and expressed and purified complexes in human cultured cells for pull-down experiments and analysis. K.J.G. performed gel-shift assays and analyzed CST-DNA binding. C.J.L., D.S.W., and T.R.C. analyzed and interpreted the model. C.J.L. and T.R.C. wrote the manuscript with input from all authors. **Competing interests:** T.R.C. is on the board of directors of Merck and a consultant for STORM Therapeutics. **Data and materials availability:** Cryo-EM maps and the de novo–built model are deposited in Electron Microscopy Data Bank (EMDB) and Protein Data Bank (PDB) with the following accession numbers: CST-3xTEL monomer (EMD-21567 and PDB ID 6W6W), CST-3xTEL decamer (EMD-21561), CST-3xTEL oligomer-mixture (Arm–EMD-21565 and head–EMD-21566), and DNA-free CST (EMD-21563). Plasmids encoding human STN1, TEN1, and CTC1 (wild-type and K1175E mutant) are available from T.R.C. under a material transfer agreement with the University of Colorado Boulder.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6495/1081/suppl/DC1
Materials and Methods
Figs. S1 to S19
Tables S1 to S3
References (*37*–*63*)

View/request a protocol for this paper from *Bio-protocol*.

23 October 2019; accepted 10 April 2020
10.1126/science.aaz9649

CHEMICAL PHYSICS

Photoelectron spectra of alkali metal–ammonia microjets: From blue electrolyte to bronze metal

Tillmann Buttersack[1,2]*, Philip E. Mason[1]*, Ryan S. McMullen[2]*, H. Christian Schewe[3]*, Tomas Martinek[1], Krystof Brezina[1,4], Martin Crhan[1], Axel Gomez[1,5], Dennis Hein[6,7], Garlef Wartner[6,7], Robert Seidel[6,7], Hebatallah Ali[3], Stephan Thürmer[8], Ondrej Marsalek[4]†, Bernd Winter[3]†, Stephen E. Bradforth[2]†, Pavel Jungwirth[1]†

Experimental studies of the electronic structure of excess electrons in liquids—archetypal quantum solutes—have been largely restricted to very dilute electron concentrations. We overcame this limitation by applying soft x-ray photoelectron spectroscopy to characterize excess electrons originating from steadily increasing amounts of alkali metals dissolved in refrigerated liquid ammonia microjets. As concentration rises, a narrow peak at ~2 electron volts, corresponding to vertical photodetachment of localized solvated electrons and dielectrons, transforms continuously into a band with a sharp Fermi edge accompanied by a plasmon peak, characteristic of delocalized metallic electrons. Through our experimental approach combined with ab initio calculations of localized electrons and dielectrons, we obtain a clear picture of the energetics and density of states of the ammoniated electrons over the gradual transition from dilute blue electrolytes to concentrated bronze metallic solutions.

Since the discovery of spectacularly colored alkali metal–ammonia solutions in the early 19th century, excess ammoniated electrons have attracted considerable attention, as reviewed recently by Zurek *et al.* (*1*) [see Thompson's classic monograph (*2*) for an overview of the older literature]. Alkali metals are soluble in liquid ammonia up to concentrations of roughly 20 mol % metal (MPM)—i.e., one metal atom per about four solvent molecules (*1*). A transition from a blue electrolyte to a bronze- or gold-colored metallic solution upon increasing alkali metal concentration is accompanied by a liquid–liquid phase separation at sufficiently low temperatures (*1*–*7*). The nature of the metallic transition in both liquid and crystalline alkali metal–ammonia systems, directly evidenced by an orders-of-magnitude increase in electrical conductivity, has puzzled researchers for decades (*1*, *8*–*10*) and is not yet understood in molecular detail. The involved chemical species include dilute solvated electrons and dielectrons as well as their various complexes with alkali metal cations (*1*)—all gradually coalescing into delocalized structures and giving rise to a conduction band. A series of conferences on this topic, the Colloques Weyl, was organized in the second half of the past century, resulting in a series of articles primarily focused on the structure, thermodynamics, and electrical and magnetic properties of the alkali metal–ammonia solutions (*11*–*16*). Electrons in liquid ammonia have also been thoroughly studied with nuclear magnetic resonance (NMR) and electron spin resonance (ESR) techniques. The latter show a narrow structureless spin resonance line with a *g* value (i.e., dimensionless magnetic moment) characteristic of a free-electron spin, which broadens upon increasing the alkali metal concentration (*2*, *17*). Shifts in the ^{1}H and ^{14}N NMR positions (Knight shifts) give a measure of the unpaired electron spin density at all constituent nuclei within the orbit of the molecules solvating the unpaired electron (*17*, *18*). Shkrob has argued (*19*) that the Knight parameters from ^{14}N NMR, electron spin echo relaxation, and ESR linewidth data can only be interpreted as a transfer of a substantial fraction of the spin density to the nitrogen atoms in the first solvent sphere.

The principal means to explore electronic structure, and thus the binding energies and density of states of excess ammoniated electrons, is photoelectron spectroscopy (PES). Liquid ammonia has a great advantage over water in that high concentrations of ammoniated electrons can be reached in solutions that are stable for extended periods of time without the danger of explosion (*20*). Nevertheless, compared with the number of such investigations of electrons solvated in water (i.e., hydrated electrons) (*21*–*23*), PES studies in liquid ammonia are scarce. Early photoelectron (PE) total emission yield experiments led to an estimate of the PE threshold of ~1.4 eV (*24*, *25*), in good agreement with electrochemical determination of the adiabatic binding energy of an ammoniated electron (*26*). This value is also roughly consistent with results from cluster extrapolations (*27*–*31*). However, clusters have only limited relevance to liquid bulk systems, as they inevitably exhibit pronounced surface effects and are typically solid rather than liquid (*32*, *33*). As a result, structures such as metastable clusters exist; these structures are characterized by low electron binding energies and have no liquid bulk analog (*33*). Electron scattering data from clusters also differ from condensed phase data (*34*). Additional insight into the ultrafast dynamics of ammoniated electrons emerged from femtosecond time-resolved experiments involving multiphoton photoionization in pure liquid ammonia or photoexcitation in dilute alkali metal–ammonia solutions (*35*–*38*). These studies have typically probed ammoniated electrons in the low-concentration regime (i.e., individual electrons well below the electrolyte-to-metal transition). In concentrated systems, plasmons in metallic lithium–ammonia solutions were explored by x-ray scattering two decades ago (*10*), and a PES study of small to medium-sized cryogenic sodium–ammonia clusters was performed recently (*31*).

There is thus a clear need for a direct PES investigation of excess ammoniated electrons that would cover both the electrolyte and metallic regimes. We have recently overcome a critical obstacle in collecting PEs from a volatile polar refrigerated liquid. We developed an experimental setup that produces a liquid ammonia microjet and performed PES measurements with this apparatus (*39*). In that study, we characterized the valence and core orbital structure of pure gaseous and liquid ammonia and quantified the effect of the condensed phase environment on the orbital energies, which was found to be even stronger than in water, despite weaker hydrogen bonding in liquid ammonia (*40*). This work has paved the way for PES investigations mapping the electrolyte-to-metal transition through the study of liquid alkali metal–ammonia solutions of increasing concentrations, as reported here.

Electronic structure calculations enable interpretation of PES measurements of ammoniated electrons in terms of a complex structural, dynamical, and molecular orbital picture. So far, only molecular pseudopotential calculations have been performed for electrons in liquid ammonia (*41*, *42*), with density functional theory (DFT) applied to crystalline alkali metal–ammonia systems (*8*). Although the early liquid-state calculations provided

[1]Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, Flemingovo nám. 2, 16610 Prague 6, Czech Republic. [2]Department of Chemistry, University of Southern California, Los Angeles, CA 90089-0482, USA. [3]Molecular Physics, Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, D-14195 Berlin, Germany. [4]Charles University, Faculty of Mathematics and Physics, Ke Karlovu 3, 121 16 Prague 2, Czech Republic. [5]Département de Chimie, École Normale Supérieure, PSL University, 75005 Paris, France. [6]Helmholtz-Zentrum Berlin für Materialien und Energie, Albert-Einstein-Strasse 15, D-12489 Berlin, Germany. [7]Department of Chemistry, Humboldt-Universität zu Berlin, Brook-Taylor-Str. 2, D-12489 Berlin, Germany. [8]Department of Chemistry, Graduate School of Science, Kyoto University, Kitashirakawa-Oiwakecho, Sakyo-Ku, Kyoto 606-8502, Japan.
*These authors contributed equally to this work.
†Corresponding author. Email: ondrej.marsalek@mff.cuni.cz (O.M.); winter@fhi-berlin.mpg.de (B.W.); stephen.bradforth@usc.edu (S.E.B.); pavel.jungwirth@uochb.cas.cz (P.J.)

some insight into the transition from individual solvated electrons through dielectrons [which exist as spin-paired singlet species in liquid ammonia (*43*)] to the onset of delocalized states upon increasing alkali metal loading, these findings were inevitably of a qualitative nature only. This was due to neglect of the explicit electronic structure of the solvent and to approximations made in the pseudopotential itself (*33*). We have shown previously that, in aqueous solutions, a quantitative picture of the electronic structure of hydrated electrons and surrounding water molecules can be obtained through DFT-based ab initio molecular dynamics (AIMD) (*44*). In this study, we employed an extension of this approach, combining it with quantum chemical embedded-cluster evaluation of electron binding energies to characterize ammoniated electrons and dielectrons.

Photoelectron spectroscopy: Electrolyte solutions

PE experiments were carried out with the SOL^{3}PES experimental setup (*45*) at the U49/2-PGM-1 beamline at the synchrotron radiation facility BESSY II (*46*) (for details, see the "Experimental Methodology" section in the supplementary materials). PE spectra at low electron binding energies of microjets of lithium–liquid ammonia solutions at alkali metal concentrations ranging from 0.012 to 9.7 MPM are presented in Fig. 1A. Analogous low-energy spectra of potassium–liquid ammonia solutions (0.15 to 1.25 MPM) and sodium–liquid ammonia (0.15 to 0.75 MPM) are shown in Fig. 1, B and C. Visually, the increase of alkali metal concentration is connected with deepening of the characteristic blue color of the solutions, with the higher concentrations becoming nearly black, even in the thin microjet, and the solution with the highest lithium concentration acquiring a discernible bronze-colored metallic sheen.

Similarly to previously studied aqueous microjets (*47*, *48*), in experiments with liquid ammonia we observed electrostatic effects leading to global PE spectral shifts. These shifts are larger for alkali metals in liquid ammonia than for solutions of alkali halide salts at equivalent concentrations (*49*). To correct for these instrumental spectral shifts, the low-concentration spectra (0.08 MPM for Li and 0.15 MPM for Na and K) were aligned horizontally such that the lowest-energy liquid ammonia peak ($3a_1$), fitted to a Gaussian function, was always anchored at 9.09 eV, which is the value of the corresponding vertical detachment energy (VDE), as determined in our recent PE measurements of a pure liquid ammonia microjet (*39*). All other spectra were aligned using the same shift. This procedure (see the supplementary materials for more details) is well justified; for low-to-moderate

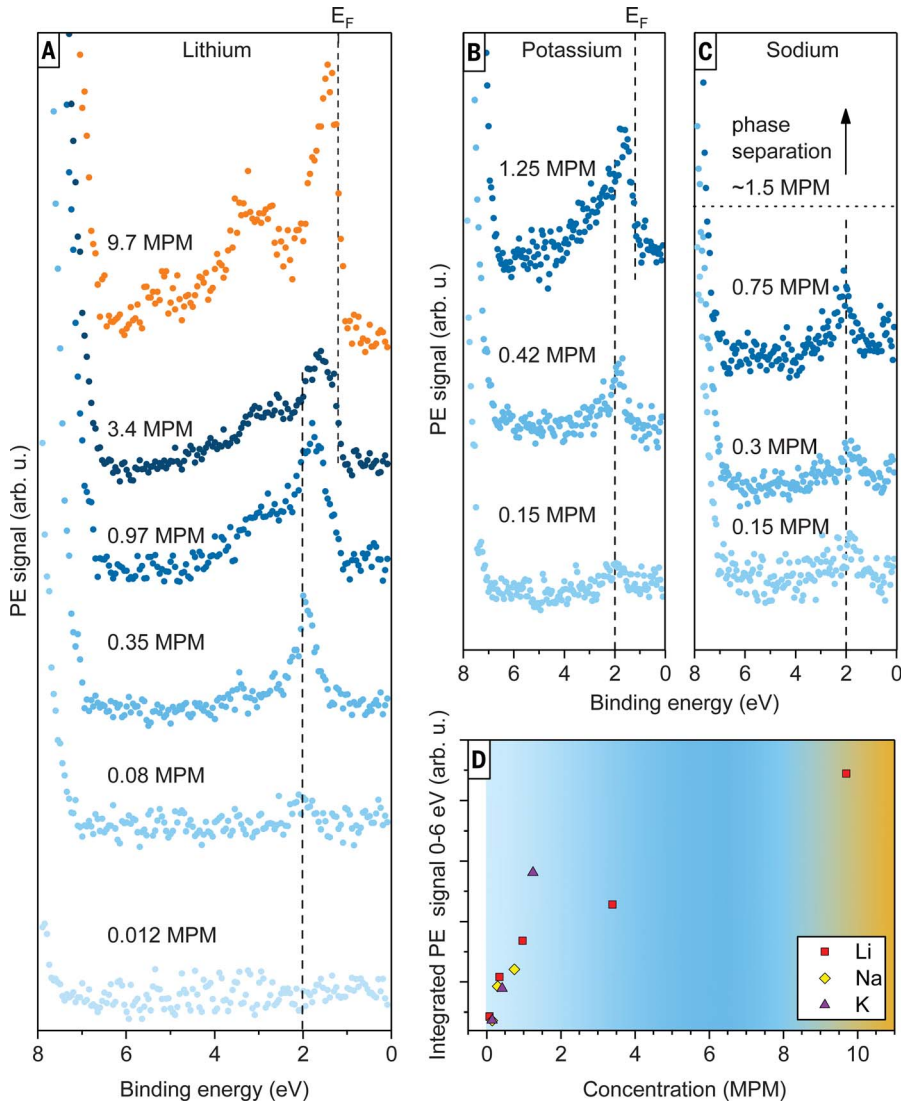


Fig. 1. Experimental PE spectra of alkali metal–liquid ammonia solutions at various concentrations, as obtained by synchrotron x-ray PES in a refrigerated liquid microjet setup. (**A**) Li in NH_3, (**B**) K in NH_3, and (**C**) Na in NH_3. Individual data points are color-coded to reflect the actual color of the solutions. Energy scales are shown with respect to the vacuum level. (**D**) Integrated peak areas from 0 to 6 eV in (A) (red), (B) (purple), and (C) (yellow), as a function of alkali metal concentration. arb. u., arbitrary units.

alkali metal concentrations, the effect of ionic solutes on the position of the solvent PE peaks has been found to be negligible in water [see below and (*47*, *48*)]. Nevertheless, these factors combine to produce a small systematic uncertainty in determining absolute values of VDEs, which we estimate not to exceed ~0.4 eV.

A notable result of the present measurements is that from ~0.08 to ~1 MPM the PE spectra consistently show a small but clearly visible peak at a VDE of ~2 eV (Fig. 1). The integrated area of this peak is roughly linearly proportional to the number concentration of the alkali metal (Fig. 1D). The observation that the position of this peak essentially does not depend on the chemical nature of the alkali metal points directly to ammoniated electrons. More precisely, starting from ~10^{-3} MPM, the solvated electrons engage in spin-pairing to form dielectrons (*1*, *2*, *50*). ESR measurements provide an estimate of the concentration dependence of the dielectron/electron ratio (*2*), which increases with dissolved metal concentration and reaches a factor of ~10 around 0.1 MPM. The measured value of ~2 eV thus corresponds primarily to the VDE of dielectrons.

Electronic structure calculations

Our experimental conclusions are further supported by electronic structure calculations. To model the structure of electrons, dielectrons,

and electron–alkali cation pairs (fig. S10) in liquid ammonia, we need to go beyond both static ab initio calculations of small clusters (*51*) and molecular pseudopotential bulk simulations (*41*–*43*). We thus combined state-of-the-art AIMD using the revPBE0-D3 hybrid density functional for sampling of relevant structures with subsequent second-order Möller-Plesset perturbation theory (MP2) for VDE calculations. The latter calculations were performed for clusters carved out of the AIMD trajectory and embedded in a polarizable continuum model (PCM). See the supplementary materials for details.

AIMD simulations of an excess electron in bulk liquid ammonia demonstrate that an ammoniated electron occupies a cavity coordinated by ~12 ammonia molecules and has a gyration radius of 3.9 Å, on average (Fig. 2A), consistent with the value of ~3.5 Å from a moment analysis of the optical absorption spectra (*52*). The spin-paired ammoniated dielectron adopts a structure similar to that of the electron (Fig. 2B), with approximately the same number of ammonia molecules in contact and a slightly larger average gyration radius of 4.4 Å. In both cases, the solvent shell is very diffuse and lacks clear separation from the rest of the solvent. Our test AIMD calculations show that adding a second electron of the same spin leads to the formation of two separate solvated electron cavities rather than a dielectron in a single cavity.

The electron solvation structure in ammonia is qualitatively similar but quantitatively different from that of a hydrated electron in water or aqueous solution (*33*, *44*). Namely, the first solvent shell of the hydrated electron is substantially more structured and less diffuse compared with those of the ammoniated electron or dielectron. Moreover, the hydrated electron is much smaller, with only four to six water molecules in its hydration shell and a gyration radius of ~2.5 Å (*44*, *52*). In regard to dielectrons, the situation in liquid ammonia is likely to be different from that in water, where hydrated dielectrons are predicted to be thermodynamically much less stable than hydrated electrons (*53*, *54*).

The AIMD simulations also serve as a basis for calculations of the VDE of the ammoniated electron and dielectron. First, we carved out the immediate electron solvation shells containing 12 NH_3 molecules from more than 100 snapshots from the AIMD trajectories. These structures were then embedded in a PCM with the dielectric constant of liquid ammonia (see below and the supplementary materials for more details). The distributions of the resulting VDEs of the two species evaluated at the MP2 level (without any additional shifts or adjustments) are plotted in Fig. 3, referenced against our experimental data. The calculated distributions have widths of ~0.3 eV, peaking at ~2.0 eV for the ammoniated electron and ~1.6 eV for the dielectron. In comparison with our low-concentration experimental spectra, we see that the experimental peak at ~2 eV encompasses within its width both the calculated solvated electron and dielectron VDE distributions (Fig. 3). These results are consistent with the previously calculated very small difference of ~0.1 eV between the lowest optical transitions of an ammoniated electron and a spin-paired dielectron in an idealized six-coordinated cluster geometry (*55*). The value for the ammoniated electron, however, differs quantitatively from extrapolations from ammonia clusters with an excess electron (*29*), yielding 1.25 eV. This is due to the fact that the cryogenic clusters are finite and solid and, therefore, have different properties from those of the bulk liquid systems described here (*56*, *57*).

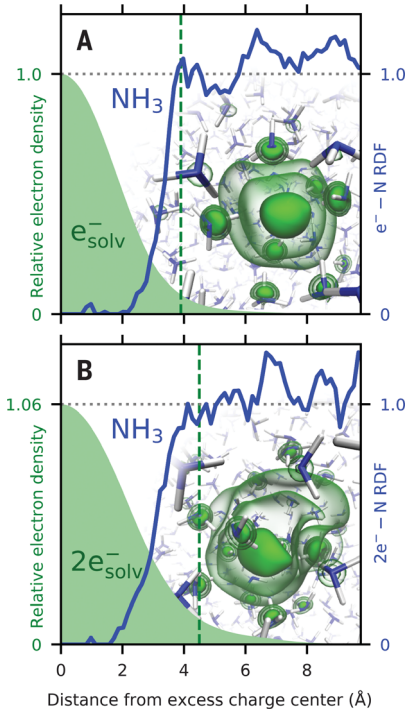


Fig. 2. Ammoniated electron and dielectron simulated by ab initio molecular dynamics (AIMD). AIMD results for (**A**) the ammoniated electron and (**B**) the ammoniated dielectron. Radial electron density profiles were calculated from the squares of the corresponding Wannier orbitals (green filled curves) and the center of excess charge—ammonia nitrogen radial distribution functions (RDF) (blue curves). Normalization is such that the integrated excess electron density of the dielectron is twice that of the electron (the latter being arbitrarily set to peak at the value of 1). Dashed vertical lines denote the electron or dielectron radius of gyration. Inset images depict the squared Wannier orbitals with surrounding ammonia molecules in the AIMD simulation box. e^-_{solv}, solvated electron.

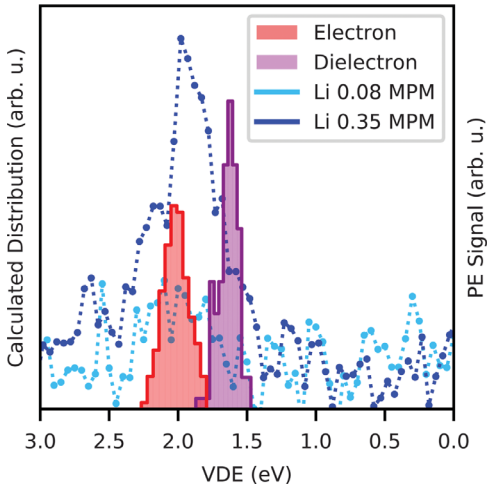


Fig. 3. Simulated VDE. An ammoniated electron (red) and dielectron (purple) were modeled using solvation shells with 12 ammonia molecules carved out from AIMD simulations and embedded in a PCM. For comparison, the corresponding experimental PE spectra of the low-concentration lithium–ammonia solutions (from Fig. 1A) are shown.

Photoelectron spectroscopy: From electrolytes to metallic solutions

Upon increasing the alkali metal concentration, the PE spectra exhibit a gradual conversion of the Gaussian-type solvated electron peak into an asymmetric band with a sharp edge toward lower binding energies accompanied

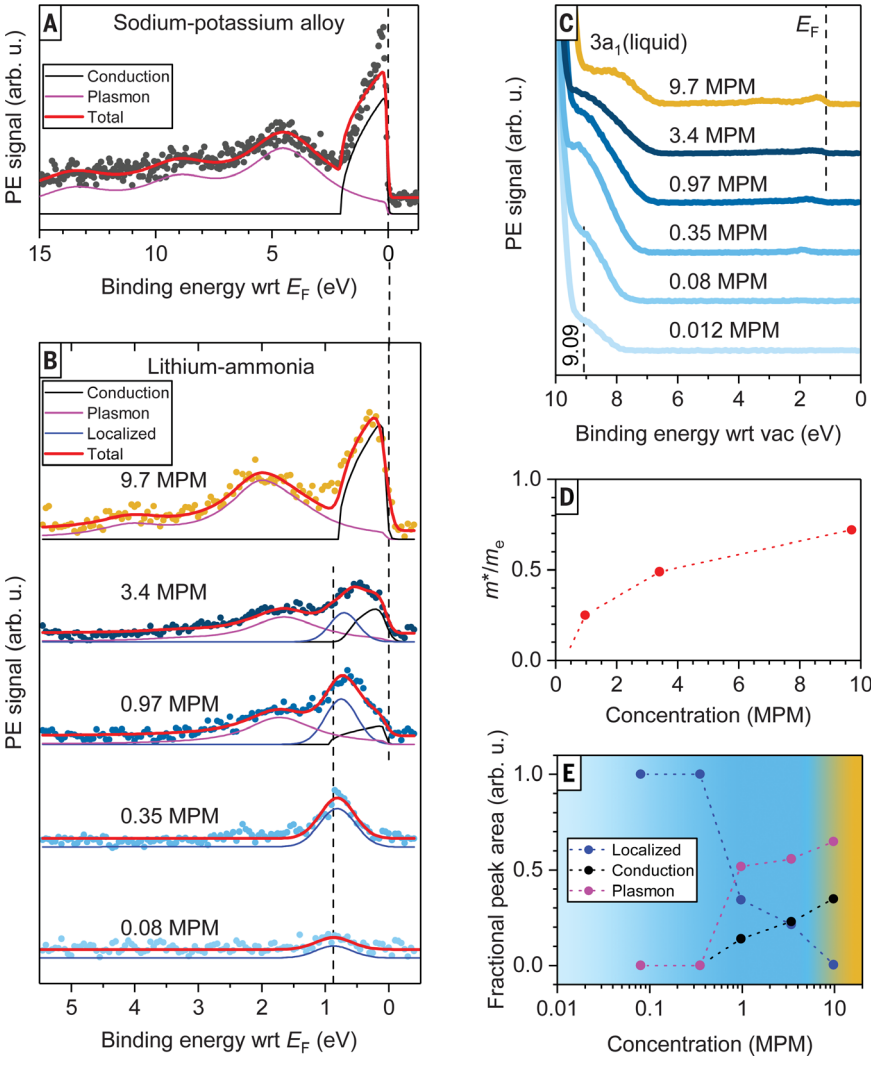


Fig. 4. Analysis and fits of the PE spectra in the electrolyte and metallic regimes. Partial fits to the conduction band, plasmons, and localized (di)electrons are vertically offset for visual clarity. (**A**) Fit (relative root mean square error of 5.3%) of the liquid Na–K alloy to a free-electron gas model. At low binding energy, we observe a Fermi edge leading feature with the characteristic parabolic shape of the conduction band; plasmon excitations are seen at higher binding energies. (**B**) Fits (relative root mean square errors of 6.3, 6.8, 7.1, 8.9, and 32.2% for 9.7, 3.4, 0.97, 0.35, and 0.08 MPM, respectively) of Li–NH_3 data to a combination, in varying ratios, of a free-electron gas model with plasmon bands for the fraction where the electron is delocalized and a single Gaussian function to represent the localized (di)electron. (**C**) Evolution of the liquid ammonia $3a_1$ peak upon increasing Li concentration. (**D**) Concentration dependence of the effective electron mass from fits in (A) and (B). (**E**) Relative peak areas corresponding to the localized Gaussian, the conduction band, and the plasmon peaks in (B). wrt, with respect to; E_F, Fermi energy; vac, vacuum; m^*, effective electron mass; m_e, stationary electron mass.

by one or two satellite peaks on the higher-binding energy side (Figs. 1 and 4, with details provided in the supplementary materials). At the lowest alkali metal concentrations, the solvated electron peak can be fitted to a Gaussian function (Fig. 4) with a full width at half maximum of about 0.45 to 0.6 eV [i.e., slightly narrower than the equivalent first ionization peak for halides in liquid ammonia (*49*)] and with a low energy onset (appearance potential) at ~1.5 eV, which is close to the previous estimate of 1.4 eV from the PE threshold measurements (*24*, *25*).

By contrast, at the highest lithium concentration of 9.7 MPM, the PE spectrum is fitted to an inverse-parabola conduction band with a sharp Fermi edge and two plasmon peaks (Fig. 4), as follows directly from the free-electron gas model for metals (*58*), with an effective electron mass close to unity (Fig. 4 and Table 1). Owing to the relatively low electron density, the plasmon frequency is in the visible range, which gives the concentrated alkali metal–ammonia solutions their characteristic bronze or gold color (*1*, *59*). A conduction band with a Fermi edge and a plasmon peak can also be observed for the 1.25 MPM potassium–ammonia solution, whereas for sodium–ammonia we could not prepare homogeneous solutions above ~1 MPM because of spontaneous phase separation at the experimental conditions (*1*, *2*).

An analogous fit to a free-electron gas model is shown for a microjet PE spectrum of liquid

Table 1. Key parameters for lithium–ammonia solutions and the sodium–potassium alloy. *c*, concentration; n_e, electron density; m_e^*, effective electron mass; m_e, stationary electron mass. Widths of the conduction band (E_c) and positions of the plasmon peak (E_P) are expressed with respect to the Fermi energy (E_F). E_c and E_P were determined from fitting to a free-electron gas model with the effective electron masses given in the table. The effective electron masses m_e^* were obtained by fitting as described in the supplementary materials. –, not determined.

Parameter	NaK	Li@NH_3	Li@NH_3	Li@NH_3	Li@NH_3	Li@NH_3	Li@NH_3
c (MPM)	100	9.7	3.4	0.97	0.35	0.08	0.012
c (M)	29	4.3	1.4	0.39	0.14	0.03	0.005
n_e ($\times 10^{21}$ cm^{-3})	16.3	2.15	1.05	0.25	0.09	0.02	0.003
m_e^*/m_e	1.21	0.72	0.49	0.25	–	–	–
E_c (eV)	2.06	0.85	0.74	0.98	–	–	–
E_P (eV)	4.52	2.03	1.67	1.74	–	–	–

50/50 sodium–potassium metal alloy (Fig. 4A and Table 1). Here, the conduction band is wider and the fundamental plasmon excitation seen at higher binding energies has a higher frequency (~4.5 eV), as expected for the higher electron density in the metal alloy compared with the metallic lithium–ammonia solutions. The higher frequency places the plasmon in the ultraviolet range when considering the optical reflectance of the sodium–potassium alloy, which does not exhibit any color and has a metallic silver sheen. This is actually true for all alkali metals except cesium, in which the lower free-electron density shifts the plasmon frequency to the visible range, conferring a golden color (*59*).

The PE spectra of the ~1 to 4 MPM lithium–liquid ammonia solutions can be fitted by a combination of a localized Gaussian at 2 eV with a conduction band and plasmon following from a free-electron gas model, albeit with a reduced effective electron mass (Fig. 4D and Table 1). As the lithium concentration increases, the relative weight of the Gaussian contribution to the spectrum decreases such that, at 9.7 MPM, it practically vanishes (Fig. 4E). At the same time, the spectra exhibit changes in the shape and position of the liquid ammonia $3a_1$ peak upon buildup of the metallic behavior of the solution (Fig. 4C). Specifically, in the electrolyte regime the position of the $3a_1$ peak almost does not change, but it does tend to broaden and move to lower binding energies upon appearance of the metallic state (for more details, see the supplementary materials).

The above results suggest that, in accord with the previous view (*1*), the electrolyte-to-metal transition upon increasing the metal concentration in alkali metal–ammonia solutions is not a sharp phase transition but rather a gradual conversion that resembles a percolation process, with an unresolved question concerning the sizes of potentially coexisting microscopic regions supporting localized and delocalized electrons (*2*, *35*, *60*, *61*). This is a different picture than that of a sharp transition at ~8 MPM drawn from recent PE spectra of alkali metal–ammonia nanodroplets (*31*). Although such experiments are pioneering in their own right, it is reasonable to question whether clusters of finite size are representative of bulk metallic solutions in their electronic structure. The cluster PE spectra exhibit notable differences, such as the lack of a sharp Fermi edge, the absence of plasmon peaks, and Fermi edge onset at higher rather than lower binding energies from the onset of the localized (di)electron peak (*31*). All of these factors suggest a qualitatively different transition in the nanodroplets, taking place at substantially higher concentrations, from those previously determined for bulk liquid systems (*1*, *31*). Our present bulk liquid PES results show a buildup of a conduction band with a Fermi edge with increasing alkali metal concentration even before the solution becomes visibly metallic (Fig. 4). This picture is also in accord with the semiquantitative Mott's criterion, which postulates that a metallic state starts to appear when the mean distance between the electrons drops below approximately four times their size (*62*). With a ~4-Å radius of gyration of the ammoniated electrons and dielectrons (Fig. 2), metallic behavior should begin to evolve at ~1 MPM, which is consistent with the onset of conduction band formation in the present PES measurements. Note, however, that the transition observed in this study is more gradual than what would strictly follow from a pure Mott's transition (*62*).

We can thus conclude that the occurrence in the PE spectrum of a conduction band with a distinct Fermi edge, together with plasmon peaks, is a signature of the electrolyte-to-metal transition. This gradual transition is observed in both the lithium–ammonia and potassium–ammonia solutions (see Figs. 1 and 4 and the supplementary materials). [As mentioned above, at concentrations exceeding ~1.5 MPM, sodium–ammonia solutions phase-separate into immiscible electrolyte and metallic phases (*2*), which compromises the microjet PE measurements.] One can also view the process from the other side—i.e., as a metal-to-electrolyte transition upon decreasing the alkali metal concentration. We see from Fig. 4 that, at the highest studied concentration of 9.7 MPM, the metallic lithium–ammonia system behaves similarly to an ideal free-electron gas, as does the liquid sodium–potassium alloy or a pure alkali metal (*63*, *64*). However, upon decreasing the concentration of the alkali metal–ammonia solutions below ~4 MPM, we observe departure from the ideal electron gas model, as exemplified by a rapid decrease of the effective electron mass well below the value of 1 m_e (where m_e is the mass of a stationary electron); see Fig. 4D. A schematic is presented in Fig. 5 to capture the essence of the transition. This image depicts the gradual interconversion between localized "chemical" species (solvated electrons and dielectrons) and delocalized "physical" moieties (metallic conduction band electrons) upon changing the electron concentration.

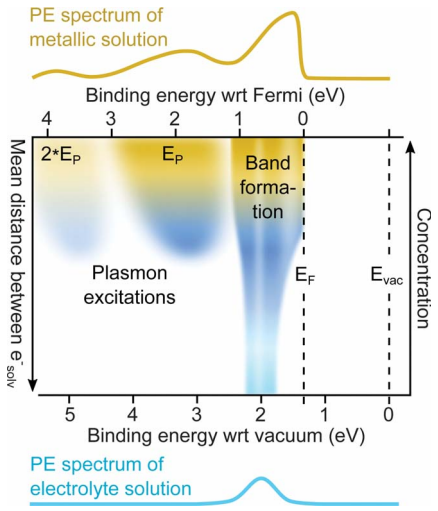


Fig. 5. Picturing the emergence of band formation during the transformation from electrolyte (blue) to metallic solution (bronze or gold) upon increasing alkali metal concentration. The limiting PE spectra are schematically represented at the top and bottom of the figure, with axes showing binding energies with respect to the Fermi and vacuum level, respectively. For a free-electron gas (upper concentration limit), the width of the leading conduction band is expected to scale with electron number density, with the plasmon bands (of energy E_P) simultaneously moving away from the leading edge.

Outlook

The present study shows that the electrolyte-to-metal transition in increasingly concentrated alkali metal–liquid ammonia solutions is a gradual process rather than an abrupt first-order transition, which is in line with previous suggestions (*1*). From the molecular point of view, this transition may be understood in a simplified way as gradual coalescence of individual solvated electrons and dielectrons upon increasing alkali metal doping, with the metallic behavior appearing around the percolation threshold.

After overcoming methodological difficulties connected to modeling the onset of the metallic state, future AIMD simulations of concentrated alkali metal–liquid ammonia solutions will shed more light on the electrolyte-to-metal transition in terms of the underlying electronic structure and molecular geometries. On the experimental side, the experience already gained from studies of liquid ammonia microjets is proving essential in our current attempts to achieve the metallic state in the much more reactive (even explosive) alkali metal–water systems.

REFERENCES AND NOTES

1. E. Zurek, P. P. Edwards, R. Hoffmann, *Angew. Chem. Int. Ed.* **48**, 8198–8232 (2009).
2. J. C. Thompson, *Electrons in Liquid Ammonia* (Clarendon Press, 1976).
3. M. T. Lodge *et al.*, *J. Phys. Chem. B* **117**, 13322–13334 (2013).
4. G. N. Chuev, P. Quémerais, *J. Chem. Phys.* **128**, 144503 (2008).
5. G. N. Chuev, P. Quémerais, J. Crain, *J. Chem. Phys.* **127**, 244501 (2007).
6. S. Hayama, N. T. Skipper, J. C. Wasse, H. Thompson, *J. Chem. Phys.* **116**, 2991–2996 (2002).
7. H. Thompson *et al.*, *J. Am. Chem. Soc.* **125**, 2572–2581 (2003).
8. J. Kohanoff, F. Buda, M. Parrinello, M. L. Klein, *Phys. Rev. Lett.* **73**, 3133–3136 (1994).
9. T. A. Kaplan, J. F. Harrison, J. L. Dye, R. Rencsok, *Phys. Rev. Lett.* **75**, 978 (1995).

10. C. A. Burns, E. D. Isaacs, P. Abbamonte, P. M. Platzman, *J. Phys. Chem. Solids* **61**, 411–413 (2000).
11. G. Lepoutre, M. J. Sienko, Eds., *Solutions Metal-Ammoniac: Proprietes Physico-Chimiques* (Benjamin, 1964).
12. J. J. Lagowski, M. J. Sienko, Eds., *Metal–Ammonia Solutions* (Butterworths, 1970).
13. J. Jortner, N. R. Kestner, Eds., *Electrons in Fluids* (Springer, 1973).
14. B. Webster, *J. Phys. Chem.* **84**, 1065–1069 (1980).
15. G. Lepoutre, *J. Phys. Chem.* **88**, 3699–3700 (1984).
16. F. Leclerq, P. Damay, *J. Phys. IV* **1**, R7 (1991).
17. P. P. Edwards, *J. Phys. Chem.* **88**, 3772–3780 (1984).
18. M. C. R. Symons, *Chem. Soc. Rev.* **5**, 337–358 (1976).
19. I. A. Shkrob, *J. Phys. Chem. A* **110**, 3967–3976 (2006).
20. P. E. Mason *et al*., *Nat. Chem.* **7**, 250–254 (2015).
21. D. M. Neumark, *Nat. Chem.* **2**, 247–248 (2010).
22. Y. Tang *et al*., *Phys. Chem. Chem. Phys.* **12**, 3653–3655 (2010).
23. F. Buchner, T. Schultz, A. Lübcke, *Phys. Chem. Chem. Phys.* **14**, 5837–5842 (2012).
24. J. Häsing, *Ann. Phys.* **429**, 509–533 (1940).
25. H. Aulich, B. Baron, P. Delahay, R. Lugo, *J. Chem. Phys.* **58**, 4439–4443 (1973).
26. T. Teherani, K. Itaya, A. J. Bard, *N. J. Chimie* **2**, 481–487 (1978).
27. H. Haberland, C. Ludewigt, H. G. Schindler, D. R. Worsnop, *Surf. Sci.* **156**, 157–164 (1985).
28. R. Takasu, F. Misaizu, K. Hashimoto, K. Fuke, *J. Phys. Chem. A* **101**, 3078–3087 (1997).
29. H. W. Sarkas, S. T. Arnold, J. G. Eaton, G. H. Lee, K. H. Bowen, *J. Chem. Phys.* **116**, 5731–5737 (2002).
30. T. Zeuch, U. Buck, *Chem. Phys. Lett.* **579**, 1–10 (2013).
31. S. Hartweg, A. H. C. West, B. L. Yoder, R. Signorell, *Angew. Chem. Int. Ed.* **55**, 12347–12350 (2016).
32. O. Marsalek, F. Uhlig, P. Jungwirth, *J. Phys. Chem. C* **114**, 20489–20495 (2010).
33. O. Marsalek, F. Uhlig, J. VandeVondele, P. Jungwirth, *Acc. Chem. Res.* **45**, 23–32 (2012).
34. T. E. Gartmann *et al*., *Phys. Chem. Chem. Phys.* **20**, 16364–16371 (2018).
35. P. Vöhringer, *Ann. Rev. Phys. Chem.* **66**, 97–118 (2015).
36. T. Vogler, P. Vöhringer, *Phys. Chem. Chem. Phys.* **20**, 25657–25665 (2018).
37. J. Lindner, A.-N. Unterreiner, P. Vöhringer, *ChemPhysChem* **7**, 363–369 (2006).
38. J. Lindner, A.-N. Unterreiner, P. Vöhringer, *J. Chem. Phys.* **129**, 064514 (2008).
39. T. Buttersack *et al*., *J. Am. Chem. Soc.* **141**, 1838–1841 (2019).
40. B. Winter *et al*., *J. Phys. Chem. A* **108**, 2625–2632 (2004).
41. Z. Deng, G. J. Martyna, M. L. Klein, *Phys. Rev. Lett.* **71**, 267–270 (1993).
42. Z. H. Deng, G. J. Martyna, M. L. Klein, *J. Chem. Phys.* **100**, 7590–7601 (1994).
43. G. J. Martyna, Z. H. Deng, M. L. Klein, *J. Chem. Phys.* **98**, 555–563 (1993).
44. F. Uhlig, O. Marsalek, P. Jungwirth, *J. Phys. Chem. Lett.* **3**, 3071–3075 (2012).
45. R. Seidel, M. N. Pohl, H. Ali, B. Winter, E. F. Aziz, *Rev. Sci. Instrum.* **88**, 073107 (2017).
46. K. J. S. Sawhney, F. Senf, W. Gudat, *Nucl. Instrum. Methods Phys. Res. A* **467–468**, 466–469 (2001).
47. B. Winter, M. Faubel, *Chem. Rev.* **106**, 1176–1211 (2006).
48. M. N. Pohl *et al*., *Chem. Sci.* **10**, 848–865 (2018).
49. T. Buttersack *et al*., *Rev. Sci. Instrum.* **91**, 043101 (2020).
50. U. Schindewolf, M. Werner, *J. Phys. Chem.* **84**, 1123–1127 (1980).
51. T. Sommerfeld, K. M. Dreux, *J. Chem. Phys.* **137**, 244302 (2012).
52. S. Golden, T. R. Tuttle, *J. Chem. Soc., Faraday Trans. II* **75**, 474–484 (1979).
53. R. E. Larsen, B. J. Schwartz, *J. Phys. Chem. B* **110**, 1006–1014 (2006).
54. K. H. Schmidt, D. M. Bartels, *Chem. Phys.* **190**, 145–152 (1995).
55. M. Mauksch, S. B. Tsogoeva, *Phys. Chem. Chem. Phys.* **20**, 27740–27744 (2018).
56. O. Marsalek, F. Uhlig, T. Frigato, B. Schmidt, P. Jungwirth, *Phys. Rev. Lett.* **105**, 043002 (2010).
57. L. Ma, K. Majer, F. Chirot, B. von Issendorff, *J. Chem. Phys.* **131**, 144303 (2009).
58. C. Kittel, *Introduction to Solid State Physics* (Wiley, 2005).
59. C. C. Addison, *The Chemistry of Liquid Alkali Metals* (Wiley, 1984).
60. J. Jortner, M. H. Cohen, *Phys. Rev. B* **13**, 1548–1568 (1976).
61. P. Damay, P. Chieux, *J. Phys. Chem.* **84**, 1203–1205 (1980).
62. P. P. Edwards, M. T. J. Lodge, F. Hensel, R. Redmer, *Philos. Trans. R. Soc. A* **368**, 941–965 (2010).
63. G. Indlekofer, P. Oelhafen, *J. Non-Cryst. Solids* **117–118**, 340–343 (1990).
64. G. Indlekofer, P. Oelhafen, *J. Non-Cryst. Solids* **156–158**, 226–231 (1993).
65. T. Buttersack *et al*., Photoelectron spectra of alkali metal-ammonia microjets. Version 1, Zenodo (2020); http://doi.org/10.5281/zenodo.3733932.

ACKNOWLEDGMENTS

S.E.B. and R.S.M. thank P. Vöhringer and P.J. thanks T. Jungwirth for valuable discussions. All authors acknowledge the Helmholtz-Zentrum Berlin for allocation of synchrotron beam time at BESSY-II. O.M., T.M., K.B., M.C., A.G., and P.J. acknowledge the Czech Ministry of Education, Youth and Sports for generous allocation of supercomputer time at IT4I in Ostrava via the Large Infrastructures for Research, Experimental Development and Innovations project IT4Innovations National Supercomputing Center – LM2015070. **Funding:** P.J. thanks the European Regional Development Fund (project ChemBioDrug no. CZ.02.1.01/0.0/0.0/16_019/0000729) for support. O.M. and K.B. were supported by the grant Primus16/SCI/27/247019 from Charles University. S.E.B. and R.S.M. are supported by the U.S. National Science Foundation (CHE-1665532). T.M. acknowledges a fellowship from the University of Chemistry and Technology, Prague, and both T.M. and K.B. received support from the IMPRS Dresden. R.S.M. acknowledges support from the States of Jersey Postgraduate Bursary. R.S. received an Emmy-Noether grant (no. SE 2253/3-1) from the Deutsche Forschungsgemeinschaft. B.W. acknowledges support from the Max-Planck-Gesellschaft. S.T. acknowledges support from JSPS KAKENHI grant no. JP18K14178. H.A. thanks the Egyptian Ministry of Higher Education and Ain Shams University for her Ph.D. grant and the Egyptian Culture Office in Berlin for support. P.E.M. acknowledges support from his YouTube popular science channel Thunderf00t. **Author contributions:** T.B., P.E.M., R.S.M., H.C.S., B.W., S.E.B., and P.J. designed the experiment and, together with R.S., H.A., D.H., G.W., and K.B., performed the measurements at the BESSY II synchrotron. S.T. assisted with data analysis. O.M., T.M., and P.J. designed the computational approach. T.M., K.B., M.C., and A.G. performed the calculations. P.J., B.W., and S.E.B. wrote the manuscript, with comments from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials or in the Zenodo repository (*65*), which is freely accessible.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6495/1086/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S28
References (*66–92*)

9 December 2019; resubmitted 25 February 2020
Accepted 3 April 2020
10.1126/science.aaz7607

WEARABLE DEVICES

Giant thermopower of ionic gelatin near room temperature

Cheng-Gong Han[1,2]*, Xin Qian[3]*, Qikai Li[1,4], Biao Deng[1], Yongbin Zhu[1], Zhijia Han[1], Wenqing Zhang[5], Weichao Wang[6], Shien-Ping Feng[4], Gang Chen[3]†, Weishu Liu[1,2]†

Harvesting heat from the environment into electricity has the potential to power Internet-of-things (IoT) sensors, freeing them from cables or batteries and thus making them especially useful for wearable devices. We demonstrate a giant positive thermopower of 17.0 millivolts per degree Kelvin in a flexible, quasi-solid-state, ionic thermoelectric material using synergistic thermodiffusion and thermogalvanic effects. The ionic thermoelectric material is a gelatin matrix modulated with ion providers (KCl, NaCl, and KNO_3) for thermodiffusion effect and a redox couple [$Fe(CN)_6^{4-}/Fe(CN)_6^{3-}$] for thermogalvanic effect. A proof-of-concept wearable device consisting of 25 unipolar elements generated more than 2 volts and a peak power of 5 microwatts using body heat. This ionic gelatin shows promise for environmental heat-to-electric energy conversion using ions as energy carriers.

The need to power Internet-of-things (IoT) sensors without using cables or batteries has spurred intense research on energy harvesting from environment. One approach is thermoelectric energy conversion technology, which is based on the Seebeck effect and uses widely available waste heat to meet the power demands of IoT sensors from microwatts to megawatts (*1*, *2*). Conventional electronic-thermoelectric (e-TE) materials are usually narrow-bandgap semiconductors that use electrons or holes as energy carriers. For a typical thermoelectric material, the thermopower (or Seebeck coefficient) is ~100 to 200 $\mu V\ K^{-1}$. As a result, generating a useful voltage of 1 to 5 V in a room temperature environment requires either the challenging integration of thousands or even tens of thousands of tiny, ~50-μm thermoelectric elements (*3*) or a DC-DC voltage booster to increase the voltage of a regular-sized device with millimeter legs up to 100 times (*4*).

An alternative route for direct energy harvesting from low-grade heat was reported in ionic systems by exploring two very different mechanisms. One mechanism is based on redox reactions at two electrodes maintained at two different temperatures. Devices using this mechanism are called thermogalvanic cells (*5*, *6*). The other mechanism is ionic thermodiffusion under a temperature gradient without redox reaction, also known as the Soret effect (*7*, *8*). Electricity can be generated continuously based on the thermogalvanic mechanism as the redox reactants are rebalanced by ionic diffusion (*9*). Thermodiffusion cells operate in a capacitive mode (*10*). After a temperature difference establishes a voltage difference, the charges stored on the electrodes can be discharged to an external load. The temperature gradient is removed for the system to recover and reapplied for next cycle. Most research is based on either the thermogalvanic or the thermodiffusion cell configuration. For thermogalvanic cells, liquid electrolytes with redox couples such as cobalt(II/III) tris(bipyridyl) (*11*, *12*), iron(II/III) (*13*), iodide/triiodide (*14*, *15*), and ferro/ferricyanide [$Fe(CN)_6^{4-}/Fe(CN)_6^{3-}$] (*9*, *16*–*23*) were reported to have an absolute temperature coefficient of a few millivolts per degree Kelvin. For example, one of the highest negative temperature coefficients of –4.2 $mV\ K^{-1}$ was realized in an aqueous system using the $Fe(CN)_6^{4-}/Fe(CN)_6^{3-}$ redox couple and chaotropic guanidium salts (*22*). For the thermodiffusion cell configuration, a thermopower of +11 $mV\ K^{-1}$ was obtained using NaOH in a polyethylene oxide (PEO) solution (*10*). Liquid cells, however, have a drawback for use in wearable devices because of the challenges of encapsulation (*24*–*26*). Quasi-solid-state electrolytes have gained attention as an alternative (*27*–*29*). A temperature coefficient of –1.09 and –1.21 $mV\ K^{-1}$ was observed when using $Fe(CN)_6^{4-}/Fe(CN)_6^{3-}$ as a redox couple in the poly(sodium acrylate) and polyvinyl alcohol matrix, respectively (*27*, *28*), a value lower than that of the redox couple in liquid solutions. High thermodiffusive thermopower is observed in the quasi-solid-state polymer gel composite of poly(vinylidene fluoride-co-hexafluoropropylene) (PVDF-HFP) and polyethylene glycol (PEG) with ionic liquid as charge carriers, and the thermopower is tunable from –4 to 14 $mV\ K^{-1}$ by tailoring the composition (*29*). Furthermore, a thermopower as high as +24 $mV\ K^{-1}$ was reported by using the high ionic selectivity of the NaOH-PEO aqueous solution in the confined nanocelluosic channels, such that Na^+ is the major mobile ion (*30*). However, whether the thermodiffusion effect and thermogalvanic effect work together synergistically to boost the final thermopower in a single ionic thermoelectric (i-TE) system remains an open question because of their fundamentally different physical pictures.

We combined thermogalvanic and thermodiffusion effects to achieve high thermopower. Before moving on, however, it is necessary to clarify our terminologies because the literature has created some confusion. Similar to conventional e-TE materials, the thermodiffusive thermopower (or Seebeck coefficient) of ions is defined as $S_{td} = -\frac{V(T_H)-V(T_C)}{T_H-T_C}$, where $V(T_H)$ and $V(T_C)$ correspond to the voltage of the hot electrode at temperature T_H and the cold electrode at temperature T_C, respectively. We clarify later that the sign of S_{td} is determined by the type of charge with higher thermal mobility in a solution, and thus is a transport property. In electrochemistry, the temperature dependence of the standard electrode potential for a reduction reaction (E^0) at the isothermal condition is referred to as the "temperature coefficient" as $\alpha_R = dE^0/dT$ (*31*, *32*), where α_R is a thermodynamic property. For a redox reaction $O + ne \leftrightharpoons R$, where the oxidized species O is converted into the reduced species R with n mole of electrons (e) transferred per unit mole of reaction, the temperature coefficient is $\alpha_R = \frac{s_R - s_O}{nF}$, where s_O and s_R are partial molar entropies of the species O and R, respectively, and F is the Faraday constant. In a thermogalvanic cell under a temperature gradient, the redox reaction contribution to the measured voltage is $V(T_H) - V(T_C) = \alpha_R (T_H - T_C)$, which means that the sign of α_R is opposite to the sign convention of the Seebeck coefficient (*33*). In addition to the redox contributions, the thermodiffusion of redox species under a temperature gradient also contributes to the total voltage, which is usually negligible (~ 10 $\mu V\ K^{-1}$) in aqueous solutions (*33*). We report a giant thermopower of 17.0 $mV\ K^{-1}$ in a quasi-solid-state i-TE material by combining the thermodiffusion effect of KCl and the temperature coefficient of a $Fe(CN)_6^{4-}/Fe(CN)_6^{3-}$ redox couple. The general strategy is to use a negative temperature coefficient (i.e., $\alpha_R < 0$) and a p-type thermodiffusive thermopower (i.e., $S_{td} > 0$) to generate a high differential thermal voltage S_i. Using such materials, a high output voltage of 2.2 V is achieved using body heat in a wearable and flexible i-TE device with only 25 unipolar elements in series working in a quasicontinuous

[1]Department of Materials Science and Engineering, Southern University of Science and Technology, Shenzhen, Guangdong 518055, China. [2]Shenzhen Engineering Research Center for Novel Electronic Information Materials and Devices, Southern University of Science and Technology, Shenzhen, Guangdong 518055, China. [3]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [4]Department of Mechanical Engineering, The University of Hong Kong, Pokfulam, Hong Kong 999077, China. [5]Department of Physics and Shenzhen Institute for Quantum Science and Technology, Southern University of Science and Technology, Shenzhen, Guangdong 518055, China. [6]Department of Electronics and Tianjin Key Laboratory of Photo-Electronic Thin Film Device and Technology, Nankai University, Tianjin 300071, China.
*These authors contributed equally to this work.
†Corresponding author. Email: liuws@sustech.edu.cn (W.L.); gchen2@mit.edu (G.C.)

mode. The proof-of-concept device demonstrates promising application of ionic gelatin in powering wearable IoT applications.

Giant thermopower of i-TE materials

We denote the as-fabricated i-TE materials as Gelatin-x MX-m/n $FeCN^{4-/3-}$ (MX = KCl, NaCl, KNO_3), where x and m/n are the molar concentrations of MX and $K_4Fe(CN)_6/K_3Fe(CN)_6$, respectively, in which $Fe(CN)_6^{4-}/Fe(CN)_6^{3-}$ serves as the redox couple (hereafter abbreviated as $FeCN^{4-/3-}$) and the ion provider MX further boosts the thermodiffusive thermopower. We chose organic gelatin for the matrix because of its abundance, low cost, high biocompatibility, and mechanical flexibility. We found that thermodiffusion of ionic species under a temperature gradient, together with the thermogalvanic effect of redox couple $FeCN^{4-/3-}$, contributes to the high thermopower of i-TE materials of Gelatin-x KCl-m/n $FeCN^{4-/3-}$. We observed an improved thermopower from 4.8 to 12.7 mV K^{-1} by increasing the concentration of KCl from x = 0 M to x = 0.8 M in the as-fabricated Gelatin-x KCl-0.42/0.25 M $FeCN^{4-/3-}$ (Fig. 1A and fig. S1). We achieved a further improved thermopower from 12.7 to 17.0 mV K^{-1} by tailoring the volume ratio of water to gelatin (Fig. 1A). This value is much higher than other reported gel-based i-TE materials by using either a thermodiffusion effect or a thermogalvanic effect (Fig. 1B and table S1).

The thermodiffusion of KCl in gelatin showed a p-type thermopower. We then used the $FeCN^{4-/3-}$ redox couple, which has a negative temperature coefficient, to achieve a synergistic effect. Because α_R is related to the entropy change of reduction reaction, the negative temperature coefficient $\alpha_R = \frac{s_{FeCN^{4-}} - s_{FeCN^{3-}}}{F} < 0$ indicates that $FeCN^{4-}$ has lower solvation entropy than $FeCN^{3-}$, which is consistent with the solvation shell being more tightly packed around $FeCN^{4-}$ because of its higher valence charge (*35*). At the hot electrode, the oxidation reaction $FeCN^{4-} \rightarrow e + FeCN^{3-}$ is thermodynamically favorable and injects electrons into the hot electrode, increasing its electrochemical potential (i.e., lower voltage) and generating a thermopower (*33*) that is consistent with the thermodiffusion contributions of KCl. At the cold side, the reduction reaction $FeCN^{3-} + e \rightarrow FeCN^{4-}$ is thermodynamically favored, with electrons attracted from the electrode, resulting in a decreased electrochemical potential (i.e., higher voltage). The redox couple therefore works together synergistically to achieve the high p-type thermopower in the as-fabricated i-TE materials of Gelatin-x KCl-m/n $FeCN^{4-/3-}$.

Optimization of thermopower

The optimization of the as-fabricated i-TE materials of Gelatin-x MX-m/n $FeCN^{4-/3-}$ involved tuning of the concentration of the ion providers (MX = KCl, KNO_3 and NaCl), the redox couple ($FeCN^{4-/3-}$), and the volume ratio of water to gelatin. We obtained a thermopower of 1.4 mV K^{-1} from $V(T_C) - V(T_H)$ and $T_H - T_C$ measurements (fig. S1) for the $FeCN^{4-/3-}$ redox couple in an aqueous electrolyte with Cu foils as the symmetric electrodes (Cu | aqueous $FeCN^{4-/3-}$ | Cu). Our measurements were in good agreement with the previously reported value (1.4 mV K^{-1}) (*9*). We observed a leap in thermopower from 1.4 to 4.8 mV K^{-1} in Gelatin-$FeCN^{4-/3-}$ (x = 0 M m/n = 0.42/0.25 M) compared with the pristine $FeCN^{4-/3-}$ solution (fig. S1). The pure gelatin had a reference thermopower of 1.3 mV K^{-1} caused by the thermodiffusion of H^+ from the ionization of carboxyl groups –COOH (*36*), whereas the

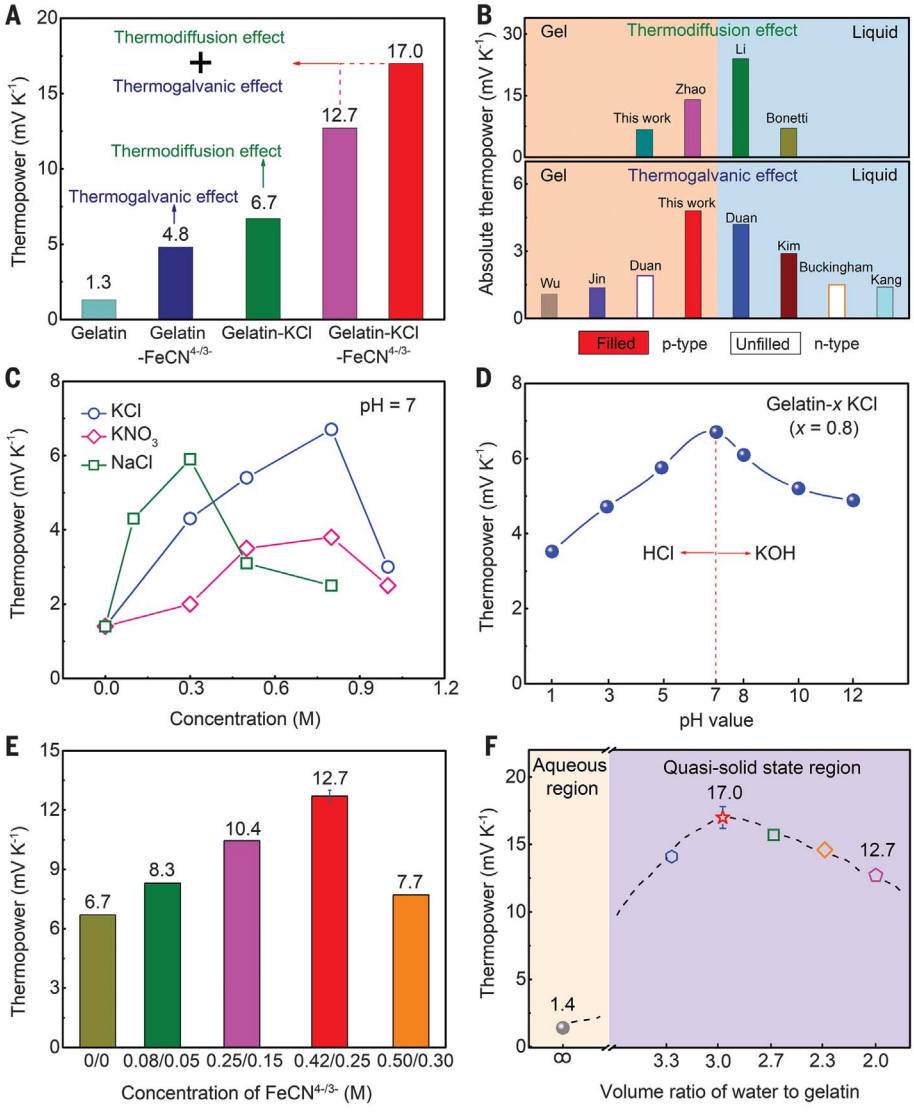


Fig. 1. Giant thermopower of i-TE materials. (**A**) Comparison of the thermopower among the as-fabricated i-TE materials of Gelatin-x KCl-m/n $FeCN^{4-/3-}$ (x is KCl and m/n are $K_4Fe(CN)_6/K_3Fe(CN)_6$ molar concentrations, respectively) in this work as Gelatin (x = 0 M, m/n = 0 M), Gelatin-$FeCN^{4-/3-}$ (x = 0 M, m/n = 0.42/0.25 M), Gelatin-KCl (x = 0.8 M, m/n = 0 M), and Gelatin-KCl-$FeCN^{4-/3-}$ (x = 0.8 M, m/n = 0.42/0.25 M, volume ratio of water to gelatin r_v = 2.0 and 3.0). (**B**) Absolute thermopower of i-TE materials containing the thermodiffusion effect or the thermogalvanic effect (table S1). The filled and unfilled columns represent p-type and n-type thermopower, respectively. (**C**) Thermopower of i-TE materials of Gelatin-x KCl, Gelatin-x KNO_3, and Gelatin-x NaCl with varying concentrations of KCl, KNO_3, and NaCl. (**D**) Thermopower of i-TE materials of Gelatin-x KCl, with varying pH values tuned by HCl and KOH, respectively. (**E**) Thermopower of i-TE materials of Gelatin-x KCl-m/n $FeCN^{4-/3-}$ with the fixed x = 0.8 M. (**F**) Thermopower with the dependence of volume ratio of water to gelatin for Gelatin-0.8 M KCl-0.42/0.25 M $FeCN^{4-/3-}$. r_v = 2.0 was maintained in (C) to (E).

Gelatin-0.42 M $FeCN^{4-}$ and Gelatin-0.25 M $FeCN^{3-}$ had a thermopower of 1.2 and 1.0 mV K^{-1} (fig. S2A), respectively.

We investigated the thermodiffusion effect of ion providers by comparing three series of gelatin-based i-TE materials: Gelatin-x KCl, Gelatin-x KNO_3, and Gelatin-x NaCl with x = 0, 0.1, 0.3, 0.5, 0.8, and 1 M, respectively (Fig. 1C and fig. S3). Gelatin-x KCl had an increased thermopower from 4.3 mV K^{-1} to a peak value of 6.7 mV K^{-1} as the concentration of KCl increased from x = 0.3 to 0.8 M, and then a decline when the concentration of KCl increased further. The Gelatin-x NaCl also had a similar peak thermopower of ~ 6.7 mV K^{-1} but at the concentration of x = 0.3 M. The Gelatin-x KNO_3 had a lower peak thermopower of ~3 to 4 mV K^{-1} in the range of x = 0.5 to 0.8 M.

Theoretically, the contribution to the thermopower of mobile cations and anions in i-TE materials could be analogous to the multiband transport in e-TE materials. The temperature gradient drove both cations and anions to migrate across the device from the hot side to the cold side, resulting in a net charge accumulation and an internal electric field that generated voltage. We derived the total thermodiffusive thermopower of a symmetrical electrolyte, such as Gelatin-x KCl, based on the Onsager transport theory as follows (*33*):

$$S_{td} = \frac{D_+\hat{S}_+ - D_-\hat{S}_-}{e(D_+ + D_-)} \quad (1)$$

where the subscript "+" or "−" denotes the ion species, e is the elementary charge, D and $\hat{S}$ are the mass diffusion coefficient and the Eastman entropy of transfer, respectively, and $\hat{S}$ is essentially the temperature dependence of the free energy dG/dT, which is related to the interaction between solutes and the surrounding media (*37*). Cations and anions in i-TE materials are equal so that the ionic thermodiffusion is ambipolar, a difference from e-TE materials. Analogous to the Einstein relation for diffusion driven by a concentration gradient, thermal mobility can also be defined as $D\hat{S}/k_BT$ (*33*). The positive thermopower suggested that the thermal mobility ($D_+\hat{S}_+/k_BT$) of cation K^+ was larger than that ($D_-\hat{S}_-/k_BT$) of anion Cl^-. The ionic interactions induced by negatively charged gelatin network could generate a larger Eastman entropy of transfer $\hat{S}_+$, which might be responsible for the large p-type thermodiffusive thermopower. Alternatively, the complicated relation between the diffusion coefficient and the concentration in the matrix with a charged polymer network may also be responsible. Experiments (*38*) and computational analysis (*39*, *40*) have shown that in a negatively charged polymer network, the cations have a higher diffusion coefficient. A small fraction of cations tend to "condensate" along negatively charged polymer chains. This "counterion condensation" was proposed by Manning (*41*). These immobilized K^+ condensed near the polymer could further impose frictional drags on Cl^-, which would reduce the mobility of Cl^-. However, the rest of K^+ not condensed around the polymer backbones remains more mobile compared with the Cl^- that was dragged by the condensed immobile K^+. We observed that the thermopower is concentration dependent. As the concentration increases, the fraction of mobile cations increases compared with the condensed cations (*39*). Further increasing the concentration could decrease the Debye length of the electrical double layer and induced a screening effect of the ionic coupling between the ions and gelatin, and the thermal mobility of ions tends to converge to pure KCl solution, which has a negligibly small thermopower measured to be ~40 μV K^{-1} (*34*). This tradeoff could explain the existence of maximum thermopower in Gelatin-x KCl and Gelatin-x KNO_3. The lower thermopower of Gelatin-x KNO_3 compared with Gelatin-x KCl can also be attributed to the smaller difference between the thermal mobility of K^+ and NO_3^-. NO_3^- is a stronger water-structure breaker compared with Cl^-, resulting in a higher mass diffusion coefficient D_- (*42*), which is consistent with the ionic conductivity measurement (fig. S4). Thus, the NO_3^- cancels more thermopower than Cl^- in the as-fabricated gelatin-based i-TE materials. Moreover, we found that the pH values affected the thermopower of the i-TE material, i.e., Gelatin-x KCl (x = 0.8 M) (Fig. 1D and fig. S5, A and B), because of the ionization of gelatin functional groups (–COOH), which could affect the ion-gelatin interaction and effectively change the Eastman entropy of transfer of ions. We observed an optimized thermopower of 6.7 mV K^{-1} at pH = 7.0. Additionally, we investigated the Gelatin-x K_2SO_4 (x = 0.25, 0.40 and 0.50 M) with divalent anions (fig. S5C). Among the investigated concentrations, the Gelatin-x K_2SO_4 (x = 0.40 M) showed the highest thermopower at 4.9 mV K^{-1}, which is much less than the 6.7 mV K^{-1} of Gelatin-x KCl (x = 0.8 M).

Adding $FeCN^{4-/3-}$ into the Gelatin-x KCl system makes the thermopower sensitive to the concentration of the $FeCN^{4-/3-}$ redox couple. The thermopower varied from 6.7 mV K^{-1} to 8.3, 10.4, 12.7, and 7.7 mV K^{-1} as x = 0.8 M, whereas m/n changed from 0/0 M to 0.08/0.05, 0.25/0.15, 0.42/0.25, and 0.50/0.30 M, respectively (Fig. 1E and fig. S2B). We repeatedly observed the highest thermopower of 12.7 mV K^{-1} in the as-fabricated i-TE material of Gelatin-0.8 M KCl-0.42/0.25 M $FeCN^{4-/3-}$ (fig. S6). We measured a lower thermopower of Gelatin-0.8 M KCl-m/n $FeCN^{4-/3-}$ with m/n = 0.25/0.25 M (11.0 mV K^{-1}) and 0.42/0.42 M (7.3 mV K^{-1}) compared with m/n = 0.42/0.25 M (fig. S2C). We attribute the high thermopower to the synergy of the thermogalvanic effect of redox couple $FeCN^{4-/3-}$ and the thermodiffusion effect of the mobile ions. Additionally, the thermal conductivity of the i-TE material Gelatin-0.8 M KCl-0.42/0.25 M $FeCN^{4-/3-}$ is low (0.15 W m^{-1} K^{-1} at 293 K), allowing it to maintain a temperature difference for power generation (fig. S7) (*33*). We observed excellent reversibility of the redox reaction evidenced by the overlapped peaks scanned for three cycles in CV curves (fig. S8). We observed the anodic and cathodic peaks from 0.05 to 0.28 V and –0.05 to –0.28 V (versus Pt), respectively, in the CV curves of the Gelatin-0.8 M KCl-m/n $FeCN^{4-/3-}$ (fig. S9A). We found increasing redox peak potential (E_p) and current density with increasing m/n values (fig. S9B). Additionally, the oxidized species ($FeCN^{3-}$) generated at the hot side and the reduced species ($FeCN^{4-}$) generated at the cold side migrated to the other electrode under a concentration gradient, making continuous current output possible (*9*, *43*).

The water/gelatin volume ratio (r_v) also boosted the thermopower of the as-fabricated Gelatin-0.8 M KCl-0.42/0.25 M $FeCN^{4-/3-}$ system. The water in the gelatin matrix provides the diffusion channel for ions in the quasi-solid-state i-TE material, affecting the thermopower (Fig. 1F and fig. S10). We varied r_v values and observed a continuous increase from 12.7 to 17.0 mV K^{-1} as r_v increased from 2.0 to 3.0. Increasing r_v further to 3.3 decreased the thermopower to 14.1 mV K^{-1} (Fig. 1F). Higher r_v also reduced the fracture strain and stretchability. We fixed r_v at 2.0 for device demonstration.

Mechanism of synergistic effect

This section explains the synergy between the thermodiffusion and thermogalvanic effects (Fig. 2, A to C). The thermodiffusion of KCl accumulated positive net charges near the cold electrode, generating an electric field pointing from the cold electrode to the hot electrode (Fig. 2A). This generated a thermodiffusive voltage $\Delta V_{td} = -\frac{\tilde{\mu}_{TH}-\tilde{\mu}_{TC}}{e} = V(T_H) - V(T_C) < 0$. The higher solvation entropy generates more $FeCN^{3-}$ than $FeCN^{4-}$ at higher T (*35*) through oxidation. This transfers electrons to the hot electrode increases the electrochemical potential ($\tilde{\mu}_{TH}$). $FeCN^{4-}$ generation was promoted and extracted electrons from the cold electrode. The T gradient drives thermodiffusion and balances the redox reaction. Consequently, the thermogalvanic effect shifts the $\tilde{\mu}$ of both electrodes in the same direction as the thermodiffusion effect. The thermogalvanic voltage that we measured was the difference in standard electrode potential $\Delta E^0 = -\frac{\tilde{\mu}_{TH}-\tilde{\mu}_{TC}}{e} < 0$, which has the same sign as the thermodiffusive voltage. The $FeCN^{4-/3-}$ also participated in thermodiffusion and contributed to the final thermopower.

From the Onsager transport formulation, we see how a large, positive thermopower comes from the coupling of the thermodiffusion and thermogalvanic effects (*33*). We observed that the total thermopower S_i could be written as a summation:

$$S_i = -\alpha_R + S_{td}(K^+ - FeCN^{4-/3-}) + S_{td}(KCl) + S_{td}(gelatin) \quad (2)$$

where $-\alpha_R$ is the contribution to thermopower from the redox reaction $FeCN^{3-} + e \leftrightharpoons FeCN^{4-}$, S_{td} is the thermopower from the thermodiffusion of mobile ions, and S_{td}(gelatin) is the intrinsic thermopower of the gelatin. We used an isothermal three-electrode system (Fig. 2D) to effectively eliminate the T gradient and determine the temperature coefficient (*33*). The contribution from the redox

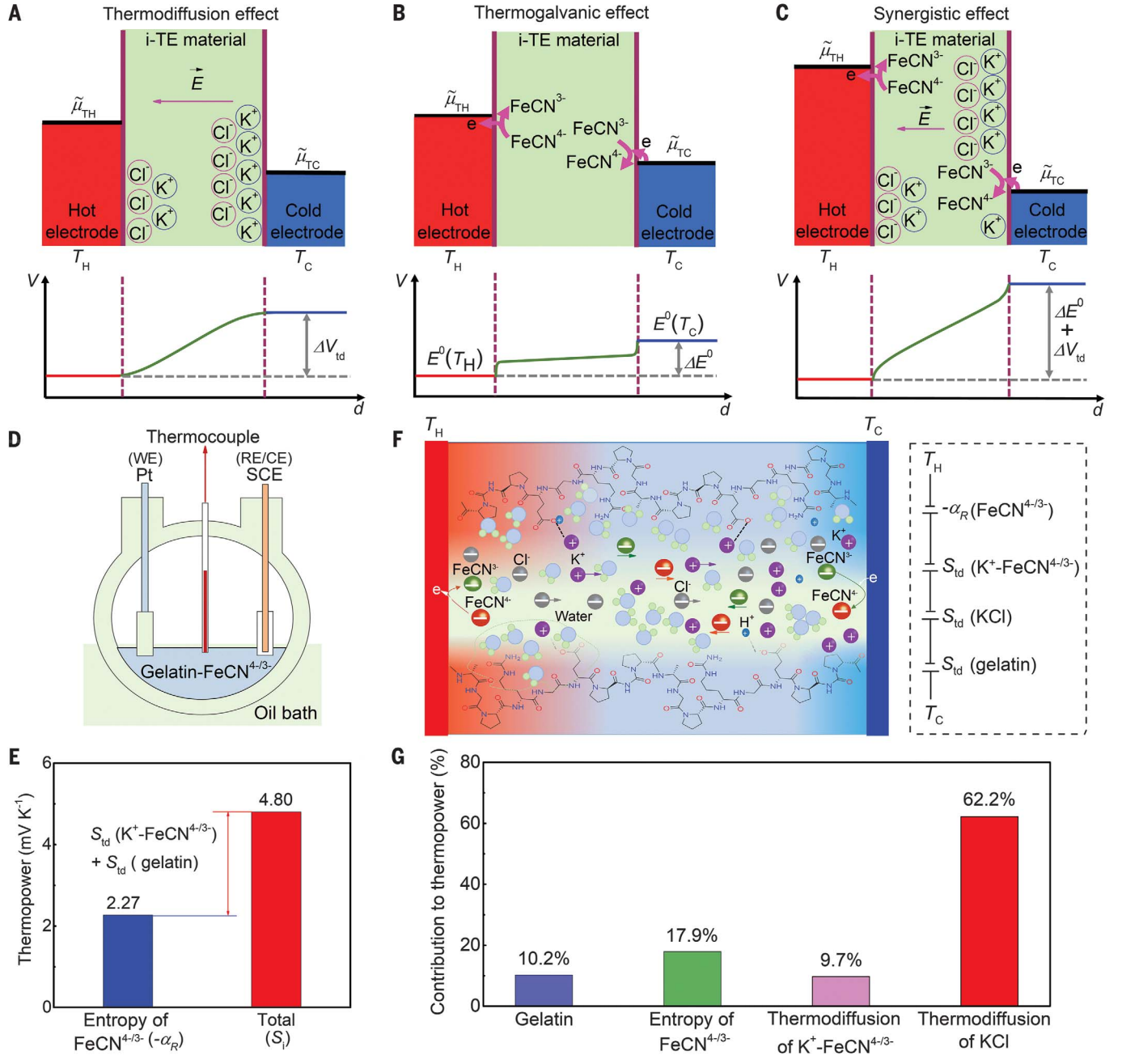


Fig. 2. Mechanism of the synergistic effect. Electrochemical potential ($\tilde{\mu}$) of charge carries diagrams and the corresponding voltage (V) distribution of i-TE material of Gelatin-x KCl-m/n $FeCN^{4-/3-}$ as (**A**) Gelatin-KCl (x = 0.8 M, m/n = 0 M), where E represents the built-in electric field; (**B**) Gelatin-$FeCN^{4-/3-}$ (x = 0 M, m/n = 0.42/0.25 M); and (**C**) Gelatin-0.8 M KCl-0.42/0.25 M $FeCN^{4-/3-}$. (**D**) Isothermal system of Gelatin-$FeCN^{4-/3-}$ for measuring the entropy of $FeCN^{4-/3-}$. The work electrode (WE) was platinum, whereas SCE was used as the reference electrode (RE) and counter electrode (CE). (**E**) Thermopower caused by redox entropy change of $FeCN^{4-/3-}$ ($-\alpha_R$) measured from (D) and total value (S_i). (**F**) Schematic figure of the diffusion, redox reaction, and interaction of the ions in the as-fabricated i-TE materials of Gelatin-x KCl-m/n $FeCN^{4-/3-}$ under the temperature gradient. (**G**) Fractional contribution to thermopower of i-TE material Gelatin-0.8 M KCl-0.42/0.25 M $FeCN^{4-/3-}$ (r_v = 2.0).

couple was finally determined to be $-\alpha_R$ = 2.27 mV K^{-1} (Fig. 2E and fig. S11) by compensating the temperature coefficient of the saturated calomel electrode (SCE) (*33*). Figure 2F shows a schematic illustration of partial contribution thermopower in a complex system containing K^+, Cl^-, $FeCN^{4-/3-}$, and water, as well as a gelatin molecule structure. Relative contribution to the total thermopower in Gelatin-0.8 M KCl-0.42/0.25 M $FeCN^{4-/3-}$ (r_v = 2.0) is determined as follows (see Fig. 2G): 10.2% contribution of Gelatin, 17.9% of redox entropy of $FeCN^{4-/3-}$, 9.7% contribution of thermodiffusion of $K_3Fe(CN)_6$ and $K_4Fe(CN)_6$, and 62.2% contribution of thermodiffusion effect of KCl (*33*). We conducted additional experiments by switching the direction of the temperature differences between two electrodes, and observed a hysteresis showing the dynamical response of the device to the transient temperature field (figs. S12 and S13) (*33*).

Working modes of an i-TE cell

A thermodiffusion effect–based i-TE cell is essentially capacitive (*10*, *30*) because the discharge current is nonfaradaic and no electrons transport across the electrode–electrolyte interfaces. A thermogalvanic cell works in a continuous manner, with redox couples reacting in opposite directions on the hot and cold electrodes and ionic diffusion supplying the reactants to electrode surfaces, thus ensuring continuous operation (*9*). We demonstrate a quasicontinuous working mode by using the i-TE material of Gelatin-0.8 M KCl-0.42/0.25 M $FeCN^{4-/3-}$ (r_v = 2.0). We assembled the i-TE cell in a laminar structure of Cu | Au | i-TE | Au | Cu (15 × 15 × 1.8 mm). We maintained the cold side at 293 K and the hot side at 301.5 K (ΔT = 8.5 K). The as-fabricated i-TE cell was charged in ~55 min to reach a high (near-saturation) voltage. We then stepped it into the quasicontinuous working mode. The cell discharged to 0 V in 10 s by connecting to an external circuit with a current linearly ramped up from 0 to maximum, and then recovered back to the high voltage in 3 min in open circuit under the same applied temperature difference. In the discharge process, the electrons flowed from the hot side to the cold side through the external circuit, resulting in a decreased internal electrostatic field and hence the cell voltage. The discharging current is also a synergistic result of redox couples and ion providers, contributed partially by the faradaic process caused by the redox couple $FeCN^{4-/3-}$ and the capacitive desorption of K^+ and Cl^-. Once the external circuit is disconnected, the diffusion of the redox couple resupplies the consumed species to the electrode and the concentration profile of ion providers reestablishes, so the cell voltage recovers, allowing for the next discharge cycle (fig. S14). We completed 100 of these charge-discharge cycles (Fig. 3A) over a time span of 5 hours. The corresponding power curve of the fifth cycle displayed parabolic behavior with the maximum at 8 μW (Fig. 3B). We expect that such quasicontinuous operation can last much longer until the electrodes are fully polarized (*33*). Output power decreased as the quasicontinuous cycle number increased (Fig. 3C, inset), which was probably caused by the polarization of the electrodes. To solve this issue, we reactivated the i-TE cell by removing the temperature difference and totally cooling down the cell while short-circuiting the electrodes. The reactivated cell recovered the voltage and current (fig. S15). The concentrations of all ionic species redistributed and the electrodes were depolarized after this process (*33*). We reproducibly achieved high thermally charged voltage over several consecutive days (fig. S16). This demonstrates that the cell can be used repetitively rather than being a one-time energy source. We reduced the thermal charge time from 3 min to ~20 s by increasing number of the layers of the i-TE cell from one to three (Cu | i-TE | Cu | i-TE | Cu | i-TE | Cu, 15 × 15 × 1.8 mm). The internal electrode shortened the time for ions to diffuse across the shortened distance, and hence shortened the thermal charging process (fig. S17).

We coated the Cu foils (10 μm thickness) with Au (40 nm) because electrode corrosion is a performance concern and found a

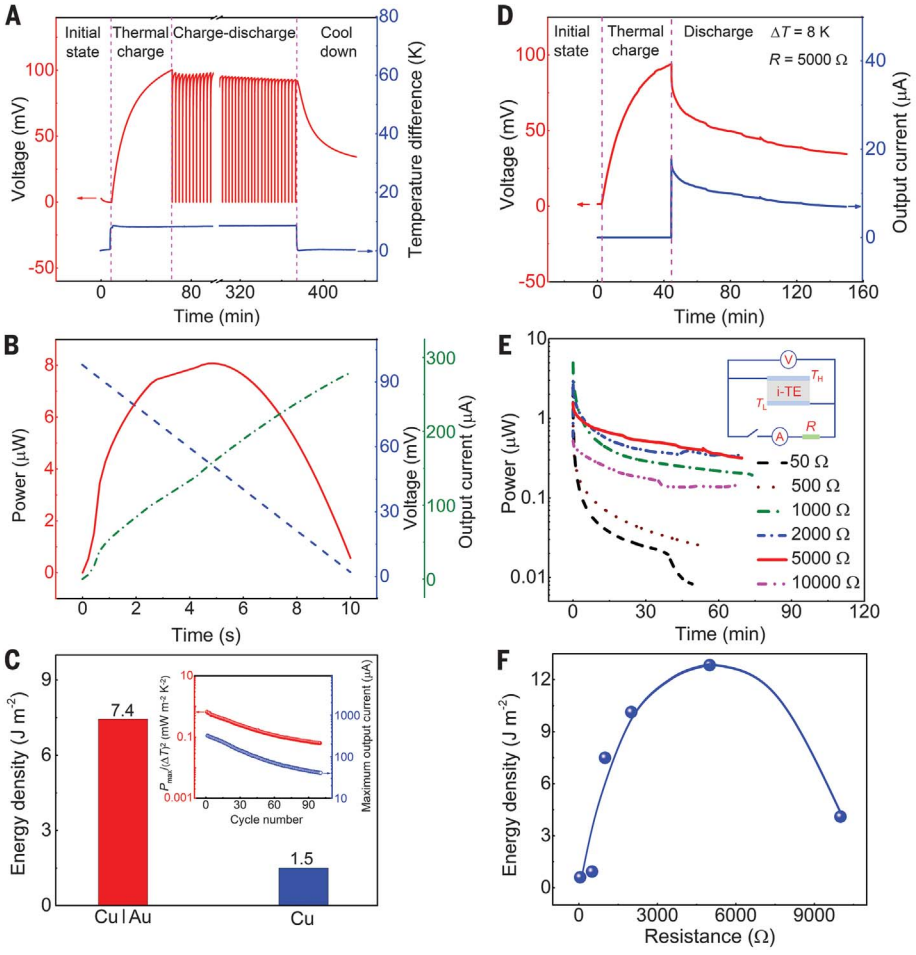


Fig. 3. Working mode of an i-TE cell. (**A**) Quasicontinuous thermal charge/electrical discharge process for an i-TE cell measured for 100 cycles [Cu | Au | i-TE | Au | Cu, 15 × 15 × 1.8 mm, Au (40 nm) coated rough Cu foils]. (**B**) Power (line), voltage (dashed line), and output current (dashed and dotted line) curves of discharge process at the fifth cycle in (A). (**C**) Corresponding total energy density of initial 50 cycles for i-TE cell with rough Cu | Au (40 nm) and smooth Cu as electrodes. Normalized output power $P_{max}/(\Delta T)^2$ and maximum output current of 100 cycles in an i-TE cell (Cu | Au | i-TE | Au | Cu, 15 × 15 × 1.8 mm) are shown in the inset. (**D**) Continuous thermal charge/electrical discharge process for the i-TE cell [Cu | Au | i-TE | Au | Cu, 15 × 15 × 1.8 mm, Au (40 nm)–coated rough Cu foils] at the external resistor R = 5000 Ω and ΔT = 8 K. (**E**) Power of the continuous discharge process at the different external resistors and ΔT ~ 8 K. The inset shows the measurement circuit. (**F**) Corresponding energy density at the different external resistors. The energy was calculated by the integration of power to time (1 hour) shown in (E).

comparable thermopower (fig. S18). However, the total energy density of initial 50 cycles was much higher (7.4 J m^{-2}) than for Cu foil electrodes (1.5 J m^{-2}) (Fig. 3C and fig. S19). The Au (40 nm)–coated Cu foil electrode has an enlarged surface area (fig. S20) (*21*, *44*). We also measured a slightly (8%) higher thermopower using a Pt electrode compared with the Cu foil electrode (fig. S21). Electrode optimization may boost the output power density of a gelatin-based i-TE cell. We calculated the specific pulsed power density, $P_{max}/(\Delta T)^2 = V_{oc}I_{sc}/(2\Delta T)^2$, where V_{oc} and I_{sc} are the open-circuit voltage and short-circuit current, respectively. We measured the maximum output power density at 0.66 mW m^{-2} K^{-2}, which is one or two orders higher than previously reported in a gel-based i-TE cell (Fig. 3C and table S2).

We show the as-fabricated i-TE cell with Au-coated copper electrodes in continuous working mode. We initially thermally charged the cell at ΔT = 8 K to reach a near-saturated voltage and then electrically discharged at the same ΔT with a constant external resistance of 5000 Ω (Fig. 3D). The output voltage and output power (Fig. 3E) initially decayed rapidly but saturated to a constant value with the external resistor, reaching steady-state thermogalvanic operation mode. We calculated the energy density (Fig. 3F) for a range of external resistance values, which had parabolic behavior and saturated at 12.8 J m^{-2}. This value is higher than that in the quasicontinuous working mode (Fig. 3C).

Proof-of-concept wearable i-TE device

An ionic liquid in a polymer gel i-TE cell based on thermodiffusion was demonstrated by Zhao *et al.* and achieved a device thermopower of 0.33 V K^{-1}. This device combined 18 pairs of n- and p-type elements (*29*). Using 25 p-type unipolar elements allowed us to reach a comparable device thermopower. Our i-TE materials are highly flexible and suitable for wearable electronics applications (fig. S22). After bending the Gelatin-0.8 M KCl-0.42/0.25 M $FeCN^{4-/3-}$ 5000 times, it had similar values of the voltage and output power density (fig. S23). The addition of KCl and $FeCN^{4-/3-}$ could potentially improve the stretchability, which we strained to 200%, compared with a 140% strain for the pure gelatin (fig. S24). The as-fabricated i-TE materials of Gelatin-0.8 M KCl-0.42/0.25 M $FeCN^{4-/3-}$ (r_v = 2.0) remained intact after stretching from 3 to 7.2 cm and recovered after release (Fig. 4A).

The giant thermopower of the as-fabricated ionic gelatin i-TE materials (Gelatin-*x* KCl-*m*/*n* $FeCN^{4-/3-}$) provides a promising solution for the voltage needed for IoT sensors in a near–room temperature environment. We constructed a flexible and wearable i-TE device assembled by serially connecting 25 i-TE elements using copper-only electrodes (Fig. 4B). This device can be worn at the back of hand (Fig. 4B, inset). We obtained a voltage of 2.2 V in a cold environment (ΔT of ~10 K). The voltage generated by our device is enough to drive different sensors without additional DC-DC

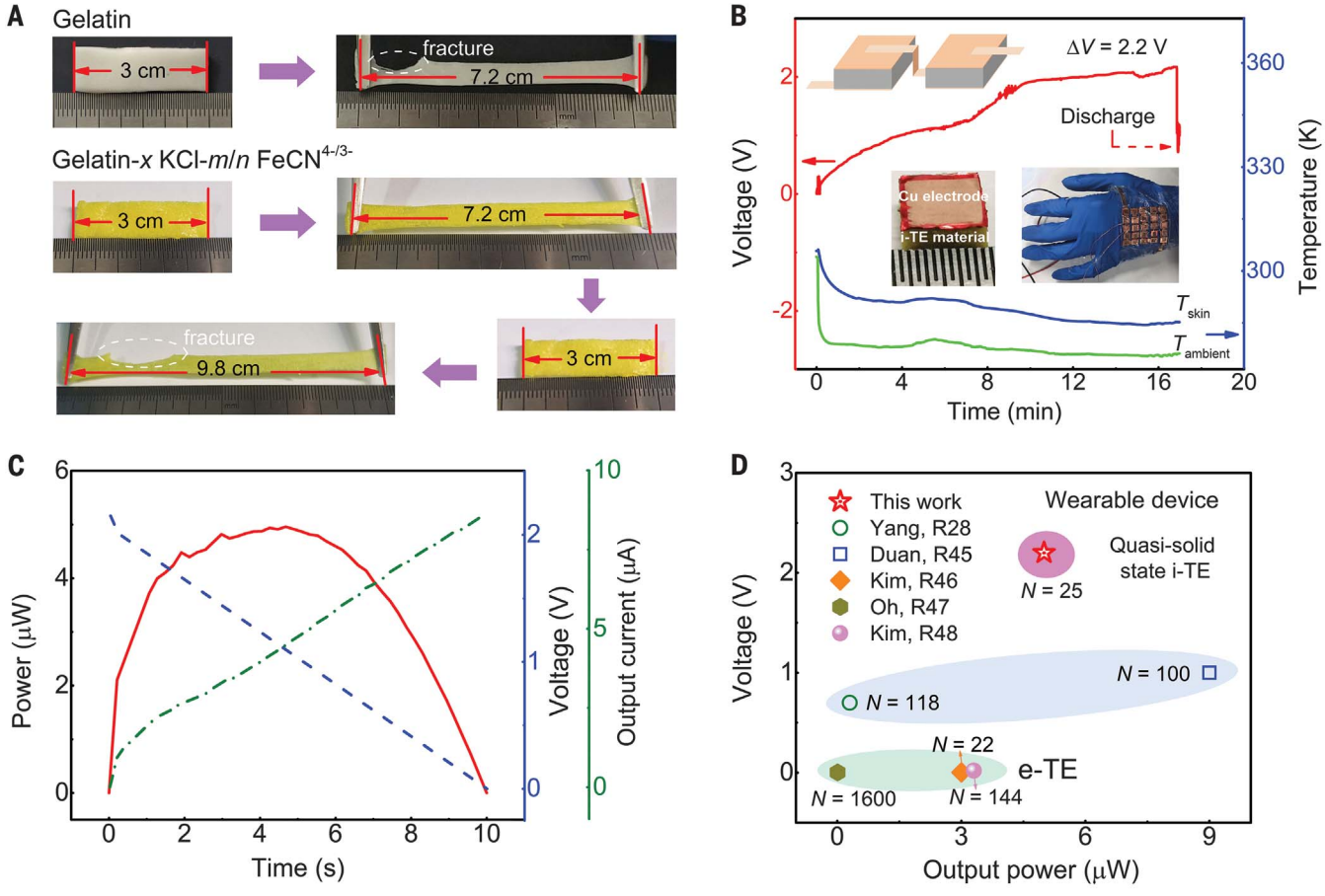


Fig. 4. Proof-of-concept of wearable i-TE device. (**A**) Tensile test of the i-TE material of Gelatin-0.8 M KCl-0.42/0.25 M $FeCN^{4-/3-}$ (r_v = 2.0) compared with pure gelatin. (**B**) Voltage generated from a proof-of-concept flexible i-TE wearable device with 25 unipolar elements (Cu | i-TE | Cu, 5 × 5 × 1.8 mm, smooth Cu foil) in series worn on the back of the human hand. (**C**) Power (line), voltage (dashed line), and output current (dashed and dotted line) curves of the proof-of-concept wearable i-TE device by harvesting real body heat. (**D**) Performance comparison in output voltage and power of the wearable device using e-TE materials and quasi-solid-state i-TE materials worn on a real human body. *N* represents the number of the n- or p-type thermoelectric elements in the wearable devices. The i-TE material used was Gelatin-0.8 M KCl-0.42/0.25 M $FeCN^{4-/3-}$ (r_v = 2.0).

voltage boosters, i.e., humidity sensors (1.6 - 3.6 V), pressure sensors (1.5 to 3.6 V), and gas sensors for monitoring indoor air quality (1.8 to 3.6 V). We measured current-voltage-output power (*I-V-P*) curves of our 25-element assembled i-TE device (Fig. 4C). We obtained a pulsed output power of 5.0 μW and a closed-circuit current of 8.5 μA in a 10-s discharge process, corresponding to the electricity energy of 3.5×10^{-5} J after a single thermal charge. This harvested energy is enough for powering many commercial sensors, i.e., 0.7×10^{-6} J of the digital temperature sensor (Si705x, Silicon Laboratories; operating voltage 1.9 to 3.6 V, 195 nA average current at 1 Hz sample rate) and 1.1×10^{-6} J of the low-power humidity sensor (HDC2010, Texas Instruments; operating voltage 1.6 to 3.6 V, 0.3 μA average current at a 1-Hz sample rate). We compared the output voltage and power of our i-TE wearable device with other reported i-TE and e-TE devices that use human body heat (Fig. 4D), and our as-fabricated i-TE wearable device was two to three times that of some previously reported i-TE devices (*28*, *45*) and two orders of magnitude higher than e-TE devices (*46*–*48*).

Summary

We have demonstrated a giant thermoelectric effect in an ionic gelatin–based i-TE material, Gelatin-x KCl-m/n $FeCN^{4-/3-}$, which synergistically combines the thermogalvanic effect of a redox couple ($FeCN^{4-/3-}$) and the thermodiffusion effect of ion providers. High positive thermopower of 12.7 ~ 17.0 mV K^{-1} was achieved by comprehensively optimizing the concentration of KCl (x = 0.8 M) and the $FeCN^{4-/3-}$ (m/n = 0.42/0.25 M) and water ratio. A proof-of-concept flexible i-TE wearable device with 25 p-type elements shows a high voltage up to 2.2 V, and a pulsed output power of 5.0 μW with total output energy of 3.5×10^{-5} J are extracted in a single discharge process by using the real heat of the human body with ΔT ~10 K, enough to power many IoT sensors. The generated voltage is two to three times higher than those of previously reported i-TE devices. The as-fabricated i-TE cell can work in a quasicontinuous thermal charge/electrical discharge mode for long-time usage, but can also work in continuous mode, delivering a maximum energy density of 12.8 J m^{-2}. This work provides a promising approach to realizing cable- and battery-free energy supplies for IoT sensors, demonstrating the promise of using ions as the energy carriers in thermoelectric energy conversion.

REFERENCES AND NOTES

1. J. He, T. M. Tritt, *Science* **357**, eaak9997 (2017).
2. M. Haras, T. Skotnicki, *Nano Energy* **54**, 461–476 (2018).
3. E. Mu *et al*., *Nano Energy* **55**, 494–500 (2019).
4. B. Iezzi, K. Ankireddy, J. Twiddy, M. D. Losego, J. S. Jur, *Appl. Energy* **208**, 758–765 (2017).
5. R. Zito Jr., *AIAA J.* **1**, 2133–2138 (1963).
6. T. I. Quickenden, Y. Mua, *J. Electrochem. Soc.* **142**, 3985–3994 (1995).
7. E. D. Eastman, *J. Am. Chem. Soc.* **50**, 283–291 (1928).
8. H. J. V. Tyrrell, D. A. Taylor, C. M. Williams, *Nature* **177**, 668–669 (1956).
9. R. Hu *et al*., *Nano Lett.* **10**, 838–846 (2010).
10. D. Zhao *et al*., *Energy Environ. Sci.* **9**, 1450–1457 (2016).
11. T. J. Abraham, D. R. MacFarlane, J. M. Pringle, *Energy Environ. Sci.* **6**, 2639–2645 (2013).
12. M. A. Lazar, D. Al-Masri, D. R. MacFarlane, J. M. Pringle, *Phys. Chem. Chem. Phys.* **18**, 1404–1410 (2016).
13. M. A. Buckingham, F. Marken, L. Aldous, *Sustain. Energy Fuels* **2**, 2717–2726 (2018).
14. T. J. Abraham, D. R. MacFarlane, J. M. Pringle, *Chem. Commun.* **47**, 6260–6262 (2011).
15. T. J. Abraham *et al*., *Electrochim. Acta* **113**, 87–93 (2013).
16. T. J. Kang *et al*., *Adv. Funct. Mater.* **22**, 477–489 (2012).
17. M. S. Romano *et al*., *Adv. Mater.* **25**, 6602–6606 (2013).
18. Y. Yang *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **111**, 17011–17016 (2014).
19. H. Im *et al*., *Nat. Commun.* **7**, 10600 (2016).
20. T. Kim *et al*., *Nano Energy* **31**, 160–167 (2017).
21. L. Zhang *et al*., *Adv. Mater.* **29**, 1605652 (2017).
22. J. Duan *et al*., *Nat. Commun.* **9**, 5146 (2018).
23. J. H. Kim *et al*., *Sci. Rep.* **9**, 8706 (2019).
24. Y. V. Kuzminskii, V. A. Zasukha, G. Y. Kuzminskaya, *J. Power Sources* **52**, 231–242 (1994).
25. S. L. Kim, H. T. Lin, C. Yu, *Adv. Energy Mater.* **6**, 1600546 (2016).
26. H. Ma *et al*., *ACS Energy Lett.* **4**, 1810–1815 (2019).
27. J. Wu, J. J. Black, L. Aldous, *Electrochim. Acta* **225**, 482–492 (2017).
28. P. Yang *et al*., *Angew. Chem. Int. Ed.* **55**, 12050–12053 (2016).
29. D. Zhao *et al*., *Nat. Commun.* **10**, 1093 (2019).
30. T. Li *et al*., *Nat. Mater.* **18**, 608–613 (2019).
31. A. J. deBethune, T. S. Licht, N. Swendeman, *J. Electrochem. Soc.* **106**, 616–625 (1959).
32. S. C. Bratsch, *J. Phys. Chem. Ref. Data* **18**, 1–21 (1989).
33. Materials and methods are available in the supplementary materials.
34. J. N. Agar, C. Y. Mou, J. L. Lin, *J. Phys. Chem.* **93**, 2079–2082 (1989).
35. B. Huang *et al*., *Phys. Chem. Chem. Phys.* **20**, 15680–15686 (2018).
36. C. M. Ofner 3rd, H. Schott, *J. Pharm. Sci.* **74**, 1317–1321 (1985).
37. S. Duhr, D. Braun, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 19678–19682 (2006).
38. J. Kamcev, D. R. Paul, G. S. Manning, B. D. Freeman, *Macromolecules* **51**, 5519–5529 (2018).
39. D. Aryal, V. Ganesan, *ACS Macro Lett.* **7**, 739–744 (2018).
40. D. Aryal, V. Ganesan, *J. Chem. Phys.* **149**, 224902 (2018).
41. G. S. Manning, *J. Phys. Chem.* **88**, 6654–6661 (1984).
42. D. J. Tobias, J. C. Hemminger, *Science* **319**, 1197–1198 (2008).
43. A. Al-zubaidi, X. Ji, J. Yu, *Sustain. Energy Fuels* **1**, 1457–1474 (2017).
44. H. A. H. Alzahrani, M. A. Buckingham, F. Marken, L. Aldous, *Electrochem. Commun.* **102**, 41–45 (2019).
45. J. Duan *et al*., *Nano Energy* **57**, 473–479 (2019).
46. S. J. Kim, J. H. We, B. J. Cho, *Energy Environ. Sci.* **7**, 1959–1965 (2014).
47. J. Y. Oh *et al*., *Energy Environ. Sci.* **9**, 1696–1705 (2016).
48. C. S. Kim *et al*., *Appl. Energy* **214**, 131–138 (2018).

ACKNOWLEDGMENTS

We thank the reviewers whose comments have helped us to improve the manuscript greatly, including identifying a sign error in the original manuscript. **Funding:** This work was supported in part by the Centers for Mechanical Engineering Research and Education at MIT and SUSTech (W.S.L. and G.C.). W.S.L. was supported by the Guangdong Innovation Research Team Project (2016ZT06G587), the Shenzhen Sci-Tech Fund (KYTDPT20181011104007), and the Tencent Foundation through the XPLORER PRIZE. W.Q.Z. acknowledges support from the Guangdong Innovation Research Team Project (2017ZT07C062), the Guangdong Provincial Key-Lab Program (2019B030301001), the Shenzhen Municipal Key-Lab Program (ZDSYS20190902092905285), and the Shenzhen Pengcheng-Scholarship Program. W.C.W. was supported by the Ministry of Industry and Information Technology of the People's Republic of China (2016YFB0901600), the Tianjin City Distinguish Young Scholar Fund, and the National Natural Science Foundation of China (21573117 and 11674289). **Author contributions:** C.-G.H. and W.L. designed the experiment, C.-G.H., Q.L., B.D., Y.Z., and Z.H. conducted the experiment. X.Q. and G.C. performed the theoretical derivation to explain the experimental data. C.-G.H., X.Q., G.C., and W.L. wrote the manuscript, W.W. and S.-P.F., W.Z., and G.C. contributed to the interpretation of the results and revision of the manuscript. All authors discussed the results and participated in revising the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6495/1091/suppl/DC1
Materials and Methods
Figs. S1 to S24
Tables S1 to S2
References (*49*–*56*)

17 September 2019; accepted 14 April 2020
Published online 30 April 2020
10.1126/science.aaz5045

REPORT

ORGANIC CHEMISTRY

Iridium-catalyzed acid-assisted asymmetric hydrogenation of oximes to hydroxylamines

Josep Mas-Roselló[1], Tomas Smejkal[2], Nicolai Cramer[1]*

Asymmetric hydrogenations are among the most practical methods for the synthesis of chiral building blocks at industrial scale. The selective reduction of an oxime to the corresponding chiral hydroxylamine derivative remains a challenging variant because of undesired cleavage of the weak nitrogen-oxygen bond. We report a robust cyclometalated iridium(III) complex bearing a chiral cyclopentadienyl ligand as an efficient catalyst for this reaction operating under highly acidic conditions. Valuable *N*-alkoxy amines can be accessed at room temperature with nondetected overreduction of the N–O bond. Catalyst turnover numbers up to 4000 and enantiomeric ratios up to 98:2 are observed. The findings serve as a blueprint for the development of metal-catalyzed enantioselective hydrogenations of challenging substrates.

Asymmetric hydrogenation with homogeneous transition metal catalysts is one of the most efficient methods for the preparation of single enantiomers at industrial scale (*1*, *2*). The enormous progress in highly enantioselective reduction of prochiral olefins and ketones is tightly linked to the development of new chiral ligand architectures (*3*, *4*). In the context of chiral amine synthesis, the catalytic asymmetric hydrogenation of imine or enamine precursors is a well-established method and is frequently used industrially (*5*). In stark contrast, related metal-catalyzed hydrogenations of oximes to produce chiral hydroxylamines have proven elusive (Fig. 1B). These substrates are often inert, and when reactivity is observed, undesired reductive cleavage of the labile N–O bond leads to primary amines (*6*). Therefore, the development of a complementary homogeneous hydrogenation mode is required.

The *N*-alkoxy amine group is an increasingly common motif in agrochemicals and pharmaceuticals, with the N–O bond offering favorable physical and biological properties (*7*). Compared to the related, more abundant, chiral amine moieties in drugs (*8*), current bioactive N–O compounds either lack chirality or are marketed as racemates (Fig. 1A). A practical asymmetric synthesis would facilitate incorporation of chiral three-dimensional hydroxylamine scaffolds as design elements in drug discovery (*9*). So far, only the use of substoichiometric to stoichiometric amounts of chiral oxazaborolidine borane adducts was shown to yield hydroxylamine products in an enantioselective fashion. However, those reactions suffer as well from undesirable primary amine by-products, depending on the oxime structure (*10*, *11*) (Fig. 1C). Moreover, costs and waste build-up make this method difficult to scale. Here, we present cyclometalated chiral iridium(III) complexes bearing a chiral cyclopentadienyl ligand (Fig. 1D, purple) and an achiral aryl imine *C*,*N*-chelate (Fig. 1D, green). We apply them for the enantioselective hydrogenation of protonated oximes to hydroxylamine derivatives, showcasing their potential in asymmetric catalysis. Related *C*,*N*-chelated half-sandwich Ir(III) complexes (*12*) have already found diverse applications in catalysis, including hydrogenation, dehydrogenation, oxidation, and hydrofunctionalization, among other transformations (*13*–*15*).

Preliminary studies revealed that cyclometalated Cp*-iridium complex **Ir1** (Fig. 2) engages in highly efficient homogeneous oxime hydrogenations in the presence of stoichiometric amounts of a strong Brønsted acid (*16*). The reaction is fully chemoselective toward reduction of the C=N bond of oxime, showing no reductive cleavage of the N–O bond. The required acid assistance in the reaction suggests an ionic hydrogenation mechanism (fig. S1), whereby a protonated substrate receives a hydride from a metal complex via an outer-sphere mechanism (*17*). The enantiodetermining facial-selective hydride delivery to the non-coordinated substrate is often the slowest step (*18*). This is distinct from classical homogeneous hydrogenation, where the substrate is bound to a metal center and subsequently receives the two hydrogen atoms from the same catalyst entity (*19*). The strong Brønsted acid fulfils a triple role: (i) The oxime substrate (~5 orders of magnitude less basic than an imine) is protonated, activating it toward hydride addition; (ii) the conjugated base dissociates from iridium, facilitating dihydrogen coordination and its subsequent heterolytic cleavage into a proton and a hydride source (*20*); and (iii) *N*-protonation of the basic hydroxyl amine product prevents catalyst poisoning (Fig. 1D). The need for stoichiometric amounts of a strong acid severely complicates the use of chiral proton sources (e.g., chiral phosphoric acids). We hypothesized that chiral cyclopentadienyl (Cp^x) ligands, which have emerged as powerful ligands for transition metal–catalyzed C-H functionalizations (*21*), constitute a potential entry point for enantioselective oxime hydrogenation. Although transient cyclometalated species of Cp^x metal complexes are frequent intermediates in C-H functionalizations (*22*), their use as stable cyclometalated complexes for catalytic purposes has been far less explored.

Oxime substrates **1** were typically obtained as *E*/*Z*-diastereomeric mixtures. When required (see below), separation by silica gel column chromatography delivered the pure benchstable *E* and *Z* isomers. Air- and moisture-stable iridium(III) complexes **Ir1** to **Ir4** were accessed in a straightforward two-step sequence. Their subsequent evaluation as selective catalysts in the reduction of oxime *E*-**1a** to *N*-*tert*-butoxylamine **2a** is summarized in Fig. 2 and fig. S2. The hydrogenation tests were conducted with 1 mol % of the iridium complex, 1.5 equivalents of methanesulfonic acid (MsOH), and 50 bar of H_2 at 23°C in 2-propanol (*16*). Exposure of *E*-**1a** to achiral complex **Ir1**, bearing an acetophenone imine as the lower chelate portion, resulted in quantitative formation of **2a** with no detected overreduction by nuclear magnetic resonance (NMR) analysis. Using (*S*)-**Ir2** where the Cp* unit was replaced by our chiral binaphthyl-derived Cp^x ligand (*23*) gave (*S*)-**2a** in 41% yield and 70:30 enantiomeric ratio (e.r.). Encouraged by this proof of principle, we tailored the lower chelating *C*,*N*-ligand architecture for the transformation resulting in **Ir3**. Additional 3,5-dimethyl groups of the aniline unit and a cyclic rigid tetralone backbone with an ethylene glycol ether adjacent to the iridium boosted the catalyst performance, giving **2a** in >99% yield and improved 89:11 e.r. In particular, the proximal (2-methoxyethyl) ether substituent rendered the complex more robust toward deactivation. In addition, the oxygen atoms of the tail might engage in hydrogen-bonding interactions with the substrate (*24*). The enantioselectivity of the hydrogenation was further improved by retaining the optimal *C*,*N*-ligand and tuning the capping chiral Cp^x ligand, resulting in **Ir4**, which has the Cp^x methoxy units replaced by phenyl groups (*25*). Using **Ir4** as the precatalyst produced **2a** in >99% yield and

[1]Ecole Polytechnique Fédérale de Lausanne (EPFL), School of Basic Sciences, Institute of Chemical Sciences and Engineering, Laboratory of Asymmetric Catalysis and Synthesis, BCH 4305, CH-1015 Lausanne, Switzerland.
[2]Syngenta Crop Protection AG, Process Chemistry Research, 4332 Stein AG, Switzerland.
*Corresponding author. Email: nicolai.cramer@epfl.ch

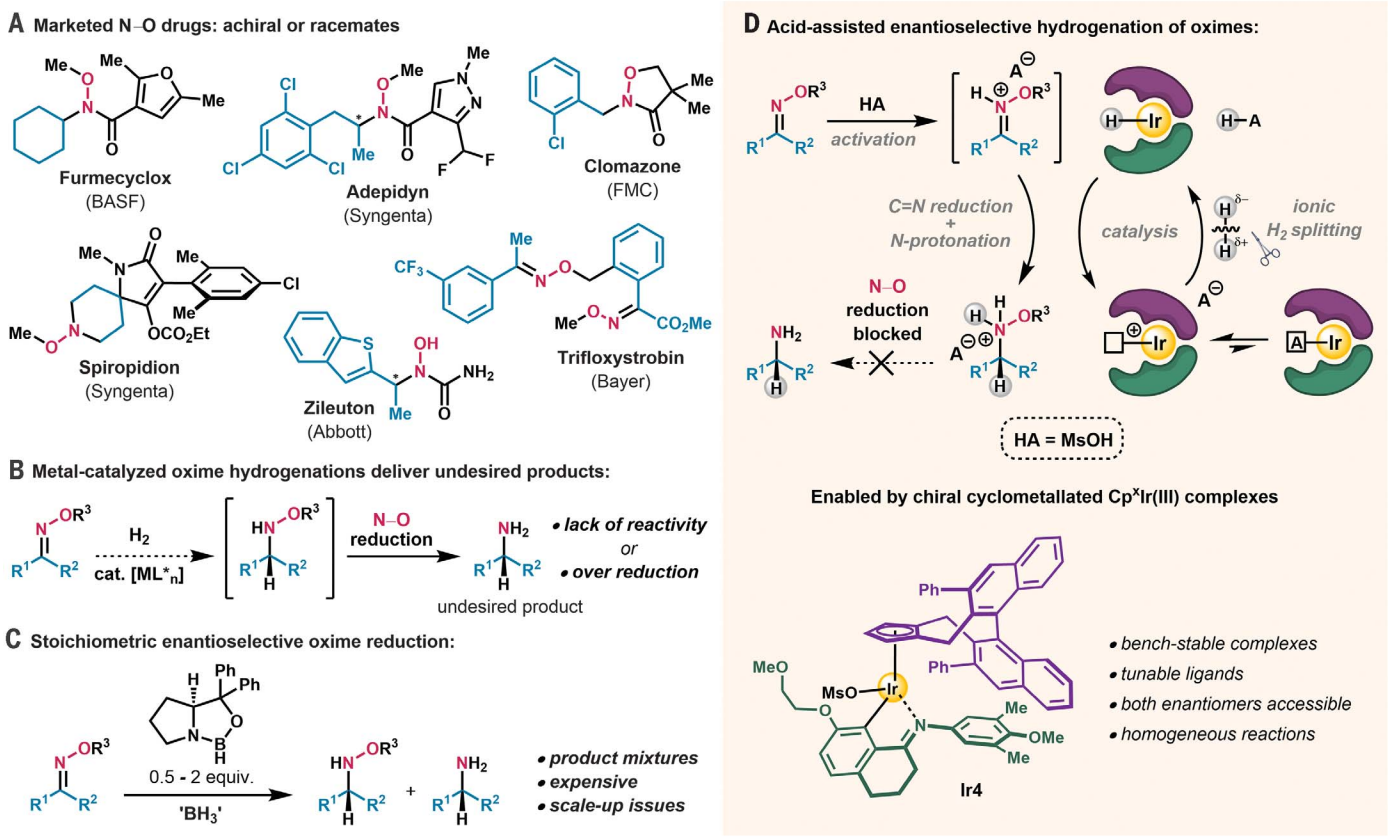
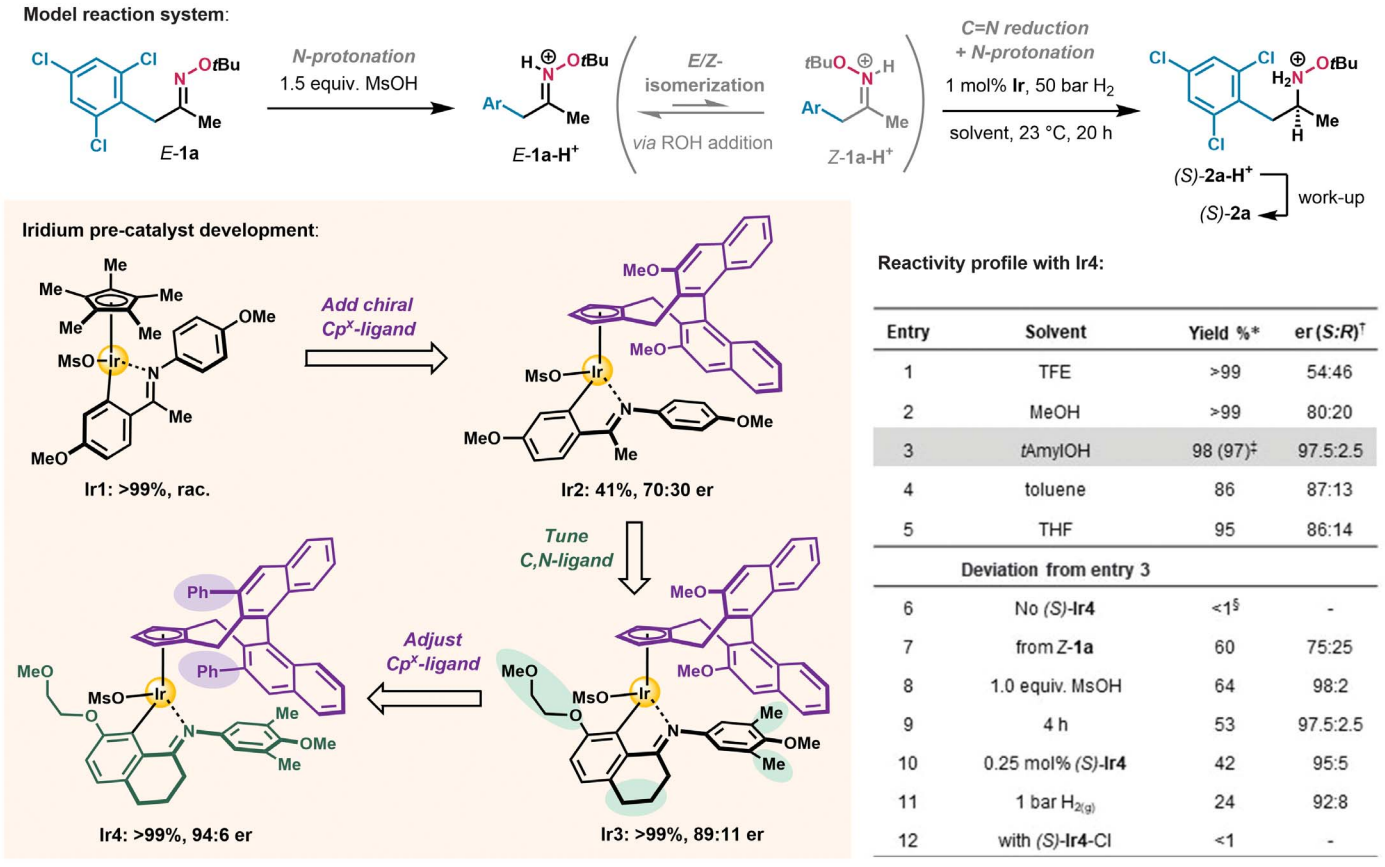


Fig. 1. Relevance of hydroxylamines and strategies for their enantioselective production by oxime reductions. (**A**) Marketed biologically active compounds containing N–O bonds either lack chirality or are sold as racemates. (**B**) Metal-catalyzed asymmetric oxime hydrogenations lead to undesired primary amine products. (**C**) Asymmetric reductions with stoichiometric chiral oxazaborolidine reagents proceed with partial cleavage of the N–O bond, are expensive, and are difficult to scale up. (**D**) Iridium(III) complexes enable the enantioselective hydroxylamine synthesis via the developed asymmetric hydrogenation platform.

94:6 e.r. ^{1}H-NMR analysis of **Ir4** in tetrahydrofuran (THF)–d_8 revealed a 70:30 diastereomeric mixture provided by the iridium stereogenic center. Upon in situ hydride incorporation, complex **Ir4-H** was detected in 96:4 diastereomeric ratio (d.r.) (fig. S3). This likely represents the ceiling of achievable selectivity. X-ray crystallographic analysis of related complex **Ir3b-I** provided additional structural insights of the complex in the solid state (fig. S4).

The hydrogenation proceeds in a variety of solvents. However, the catalyst reactivity and selectivity were lower in aprotic solvents such as toluene and THF (Fig. 2, entries 4 and 5). Among alcoholic solvents, *tert*-amyl alcohol (*t*AmylOH) was superior, giving excellent levels of enantioselectivity (97.5:2.5 e.r.) and conversion (98%) for the reduction of *E*-**1a** to **2a** (entry 3), whereas the reduction of *Z*-**1a** under otherwise identical conditions resulted in a modest conversion and modest enantioselectivity (entry 7). In line with previous reports (*26*), this indicates a strong impact of the oxime *E*/*Z* stereochemistry on the reaction outcome, with *E*-**1a** providing superior reactivity and selectivity. Because protonated oximes can isomerize by nucleophilic addition/elimination to the C=N double bond, the choice of the alcoholic solvent and the Brønsted acid is key to achieving high selectivity by tuning the substrate *E*/*Z* equilibration rate versus the hydrogenation rate. Pure *E*-**1a** isomer equilibrates to an 80:20 *E*/*Z*-mixture in the absence of **Ir4** (entry 6). Using the catalyst in *t*AmylOH produced **2a** in 97.5:2.5 e.r., irrespective of the reaction time (4 or 20 hours); this indicates that the hydrogenation is faster than oxime isomerization (entries 3 and 9). Performing the reduction in methanol (MeOH), which triggers a faster oxime isomerization, caused a drop in selectivity to 80:20 e.r. (entry 2). In contrast, 1.0 equivalents of MsOH, the minimum required amount of acid, in *t*AmylOH afforded **2a** in 98:2 e.r., although this was accompanied by incomplete conversion of *E*-**1a** (entry 8). 2,2,2-Trifluoroethanol disrupted essential interactions for stereocontrol and yielded **2a** in modest 54:46 e.r. (entry 1). Additional experiments showing the impact of the acid strength and stoichiometry on the reaction outcome are summarized in fig. S4. A high enantioselectivity was maintained at 0.25 mol % catalyst loading (entry 10) or with lower hydrogen pressure (1 bar H_2) (entry 11). With lower loadings, the risk of catalyst deactivation by chloride anion contamination increases. The chloride analog of (*S*)-**Ir4** was completely catalytically incompetent as a result of the high Cl-Ir binding affinity (entry 12). The absolute configuration of **2a** was confirmed by single-crystal x-ray analysis of its 4-nitrobenzenesulfonic acid salt.

Using the optimized conditions with a slow *E*/*Z*-oxime equilibration regime, the acid-assisted enantioselective reduction was applicable to a variety of oximes **1**, producing the corresponding alkoxy amines **2** in excellent yields and enantioselectivity (Fig. 3). Contrasting the reported racemic $B(C_6F_5)_3$-catalyzed oxime hydrogenation (*27*), the bulky *tert*-butoxy group of **1a** is not mandatory. Substrates with *O*-methyl as well as other primary and secondary *O*-alkyl substituents were quantitatively hydrogenated in high enantioselectivities, up to 97:3 e.r. Even free oximes were smoothly reduced to the corresponding hydroxylamine products without any detectable N–O bond



Entry	Solvent	Yield %*	er (S:R)†
1	TFE	>99	54:46
2	MeOH	>99	80:20
3	*t*AmylOH	98 (97)‡	97.5:2.5
4	toluene	86	87:13
5	THF	95	86:14
	Deviation from entry 3		
6	No (*S*)-**Ir4**	<1§	-
7	from *Z*-**1a**	60	75:25
8	1.0 equiv. MsOH	64	98:2
9	4 h	53	97.5:2.5
10	0.25 mol% (*S*)-**Ir4**	42	95:5
11	1 bar $H_{2(g)}$	24	92:8
12	with (*S*)-**Ir4**-Cl	<1	-

Fig. 2. Development of the iridium precatalyst and reaction optimization. Conditions: *E*-**1a**, 1 mol % **Ir**, 50 bar H_2, 1.5 equiv MsOH, 0.5 M in *i*PrOH, 23°C, 20 hours, basic work-up to free **2a**. *Determined by ^{1}H-NMR with internal standard; yield of **2a** virtually identical to the consumption of **1a**. †e.r. determined by chiral HPLC. ‡Isolated yield. §>99% of **1a** recovered (4:1 *E*:*Z*).

scission, as exemplified by **2b**. Despite the acidic reaction conditions, an acid labile acetal moiety (**2c**) and *O*-benzyl group susceptible to hydrogenolysis (**2d**, **2e**) did not interfere and remained intact. Substrate selectivity control by existing stereocenters at the ether oxygen atom, as exemplified by **1e** with the (*S*)-α-methylbenzyl group, could be overridden by the iridium catalyst. Product **2e** was obtained either in a 4:1 d.r. for the mismatched case with (*S*)-**Ir4** or in a 7:93 d.r. using the matched catalyst isomer (*R*)-**Ir4**. In general, the enantiomeric purity of products **2** could be upgraded by simple recrystallization of their salts, as shown for compound (*S*)-**2g** (from 93:7 e.r. to 99:1 e.r. with 81% recovery). When starting from pure *Z*-**1g** isomer, the antipode (*R*)-**2g** was formed in 20:80 e.r. Dialkyl oximes **1l** and **1m** bearing 2° and 3° alkyl substituents were equally well reduced, giving for instance *N*-**2m**, the *N*-methoxy derivative of rimantadine (*28*) in 93:7 e.r. Related substrates with adjacent hydroxyl or tosylate groups at the α-carbon atom were compatible, affording reduced products **2n** and **2o** in excellent yields and high selectivities. Another salient application is the access to *N*-alkoxy amino acid derivatives, as shown for valine analog **2p** formed in 98% yield and 90:10 e.r. The protocol works with D_2 generated by Skrydstrup's two-chamber setup, thus qualifying for practical deuterium isotopic labeling of valuable hydroxylamine scaffolds. Accordingly, benzoxazine **2q**-d_1 was synthesized in 80% yield and 97:3 e.r. with 92% deuterium incorporation.

We next targeted the synthesis of *N*-alkoxy derivatives of valuable chiral α-branched benzylamines (*29*). No stereocontrol was achieved with 4-methoxy acetophenone derived oxime *E*-**1j**, which equilibrates to a 9:1 *E*:*Z* isomeric mixture under the reaction conditions. Sterically more demanding *Z*-**1k**, with a locked Z-configuration due to the *tert*-butyl substituent, gave **2k** in 92:8 e.r. To further investigate the impact of the aryl oxime stereochemistry on the reaction, we individually subjected the separated isomers *Z*-**1rc** and *E*-**1rc** to conditions for fast hydrogenation (3 mol % **Ir4**) and slow isomerization (*i*PrOH, 1.0 equiv of MsOH, 2 hours) (Fig. 3B). The diastereoisomer having the *N*-OR moiety trans to the large substituent (here *Z*-**1rc**) reacted much faster and more selectively to **2rc** (99%, 97:3 e.r.) than did *E*-**1rc** (22%, 70:30 e.r.), which suggests that actually only ~7% of isomer *E*-**1rc** was reduced. Given this behavior, we hypothesized that using conditions with a fast *E*/*Z* equilibration regime would be most beneficial. Indeed, hydrogenation of a 1:1 *E*:*Z* mixture of **1rc** under equilibrating conditions [ethanol (EtOH), 1.5 equiv of MsOH] formed **2rc** in quantitative yield and 92:8 e.r.

A variety of substrates **1ra** to **1rk** were conveniently hydrogenated as the unresolved 1:1 *E*/*Z* mixtures, forming the corresponding *N*-methoxy amines in excellent yields and comparably high enantioselectivities. This substrate type allowed the demonstration of the unique chemoselectivity and functional group tolerance of the acid-assisted reduction method. Whereas most transition metal–catalyzed hydrogenation methods frequently reduce aromatic bromo, vinyl, and nitro as well as azido groups, these remarkably remained untouched under our reaction conditions. Moreover, the pinacol boronate group of **2rk** survived, serving as a potential handle for subsequent cross-coupling reactions. Nitrogen and sulfur heteroarenes **2re** and **2rf** were also compatible. (See fig. S6 for tests of additional functional

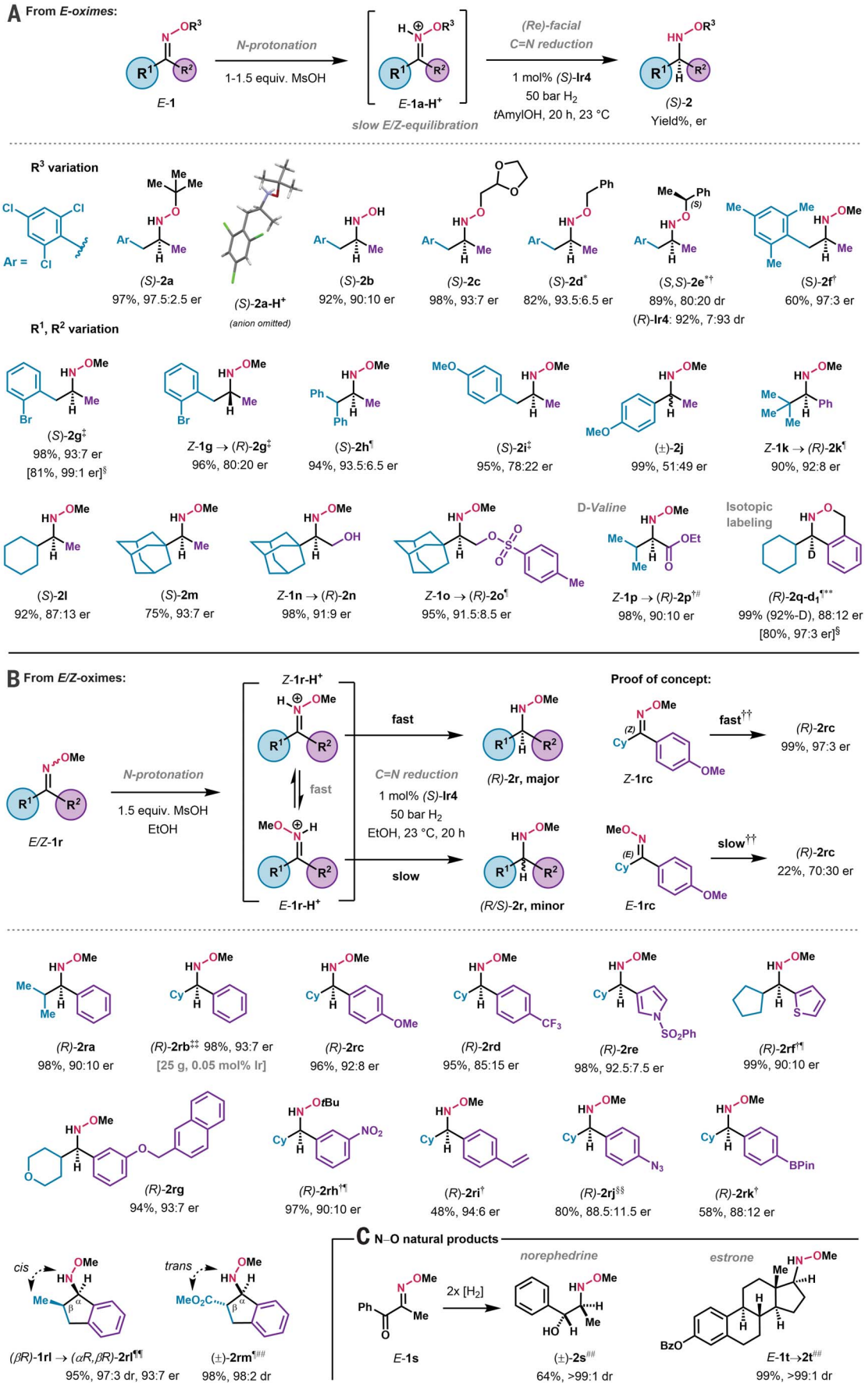


Fig. 3. Substrate scope of the oxime hydrogenation. Isolated yields; e.r. determined by chiral HPLC. (**A**) Hydrogenation of pure *E*-oximes. (**B**) Hydrogenation of *E*/*Z* oxime mixtures. (**C**) Natural product hydroxylamine derivatives. *In *i*PrOH. †2 mol % **Ir4**. ‡5.0 equiv of TFA instead of MsOH. §Recrystallized as its *p*-nitrobenzenesulfonate. ¶In MeOH. #In EtOH. **Ex situ generation of D_2 using COware. ††3 mol % **Ir4**, 1.0 equiv MsOH, *i*PrOH, 23°C, 2 hours. ‡‡94:6 e.r. at 1 mol % **Ir4**. §§20 bar H_2, 5 hours. ¶¶From (βS)-**1rl**: (αS,βS)-**2rl** 99%, 77:23 d.r. ##1 mol % **Ir1**.

groups.) Enzyme inhibitor **1rg** (*30*) could be reduced in quantitative yield and 93:7 e.r., illustrating the applicability of the method for late-stage derivatization. Existing stereogenic centers adjacent to the C=N bond exhibit a strong substrate control. For instance, (*R*)-**1rl** possessing a β-methyl stereogenic center yielded cis-isomer (*R*,*R*)-**2rl** in 97:3 d.r. and 93:7 e.r. In contrast, β-ester analog (±)-**1rm** gave trans-isomer **2rm** in 98:2 d.r., likely by cis-hydrogenation followed by acid-promoted epimerization of the ester to the thermodynamically more stable product. *N*-methoxy derivatives of norephedrine **2s** and estrone **2t** were formed in highly diastereoselective fashion, the former after double hydrogenation of **1s**. Finally, a 25-g batch of *E*/*Z*-**1rb** was hydrogenated with a reduced 0.05 mol % catalyst loading to (*R*)-**2rb** in a quantitative manner (turnover number 4000) in 93:7 e.r., supporting the scalability of the method. Overall, our findings serve as a blueprint for further asymmetric transition metal–catalyzed ionic hydrogenations of challenging substrates, and highlight the yet untapped potential of cyclometalated chiral Cp^x metal complexes in catalysis.

REFERENCES AND NOTES

1. D. J. Ager, A. H. M. de Vries, J. G. de Vries, *Chem. Soc. Rev.* **41**, 3340–3380 (2012).
2. H. U. Blaser, E. Schmidt, Eds., *Asymmetric Catalysis on Industrial Scale: Challenges, Approaches and Solutions* (Wiley-VCH, 2010).
3. P. G. Andersson, I. J. Munslow, Eds., *Modern Reduction Methods* (Wiley-VCH, 2008).
4. D. Wang, D. Astruc, *Chem. Rev.* **115**, 6621–6686 (2015).
5. J.-H. Xie, S.-F. Zhu, Q.-L. Zhou, *Chem. Rev.* **111**, 1713–1760 (2011).
6. A. M. Maj, I. Suisse, F. Agbossou-Niedercorn, *Tetrahedron Asymmetry* **27**, 268–273 (2016).
7. Z. Rappoport, J. F. Lebman, Eds., *The Chemistry of Hydroxylamines, Oximes and Hydroxamic Acids* (Wiley, 2008), chap. 13.
8. T. C. Nugent, Ed., *Chiral Amine Synthesis: Methods, Developments and Applications* (Wiley-VCH, 2010).
9. D. C. Blakemore *et al*., *Nat. Chem.* **10**, 383–394 (2018).
10. M. P. Krzemiński, M. Zaidlewicz, *Tetrahedron Asymmetry* **14**, 1463–1466 (2003).
11. R. Dumeunier, J. Kessabi, S. V. Wendeborn, H. Nussbaumer, international patent WO2014206855 (2014).
12. D. L. Davies *et al*., *Dalton Trans.* **2003**, 4132–4138 (2003).
13. C. Wang, J. Xiao, *Chem. Commun.* **53**, 3399–3411 (2017).
14. C. Michon, K. MacIntyre, Y. Corre, F. Agbossou-Niedercorn, *ChemCatChem* **8**, 1755–1762 (2016).
15. G. Zhou *et al*., *ACS Catal.* **8**, 8020–8026 (2018).
16. T. Smejkal, international patent WO2020094527A1 (2020).
17. O. Eisenstein, R. H. Crabtree, *New J. Chem.* **37**, 21–27 (2013).
18. M. P. Magee, J. R. Norton, *J. Am. Chem. Soc.* **123**, 1778–1779 (2001).
19. A. Comas-Vives, G. Ujaque, A. Lledós, *Adv. Inorg. Chem.* **62**, 231–260 (2010).
20. Z. M. Heiden, T. B. Rauchfuss, *J. Am. Chem. Soc.* **131**, 3593–3600 (2009).
21. C. G. Newton, D. Kossler, N. Cramer, *J. Am. Chem. Soc.* **138**, 3935–3941 (2016).
22. C. G. Newton, S.-G. Wang, C. C. Oliveira, N. Cramer, *Chem. Rev.* **117**, 8908–8976 (2017).
23. B. Ye, N. Cramer, *J. Am. Chem. Soc.* **135**, 636–639 (2013).
24. P. A. Dub, J. C. Gordon, *ACS Catal.* **7**, 6635–6655 (2017).
25. D. Kossler, N. Cramer, *J. Am. Chem. Soc.* **137**, 12478–12481 (2015).
26. A. S. C. Chan *et al*., *J. Chem. Soc. Chem. Commun.* **1995**, 1767–1768 (1995).
27. J. Mohr, M. Oestreich, *Angew. Chem. Int. Ed.* **53**, 13278–13281 (2014).
28. T. P. Stockdale, C. M. Williams, *Chem. Soc. Rev.* **44**, 7737–7763 (2015).
29. M. Breuer *et al*., *Angew. Chem. Int. Ed.* **43**, 788–824 (2004).
30. J. F. Dellaria, L. J. Dorn, D. W. Brooks, Abbott Laboratories, U.S. patent 5354865A (1994).

ACKNOWLEDGMENTS

We thank R. Scopelliti and F. Fadaei Tirani for x-ray crystallographic analysis of compounds **Ir3b-I** and (*S*)-**2a-H$^+$**, C. J. Cope and S. Masala for technical support, and V. Grushin and A. Robinson for insightful discussions. **Funding:** Supported by Syngenta Crop Protection AG and EPFL. **Author contributions:** T.S. conceived the project. N.C. conceived and supervised the project. J.M.-R. performed the experimental work. All authors designed the experiments, analyzed the data, and drafted the manuscript. **Competing interests:** The authors have filed a patent application (GB 1818117.2) for the described iridium-catalyzed asymmetric ionic oxime hydrogenation. **Data and materials availability:** Crystallographic data for compounds **Ir3b-I** and (*S*)-**2a-H$^+$** are available free of charge from Cambridge Crystallographic Data Centre under reference numbers CCDC 1973002 and 1973003, respectively. All other characterization data and detailed experimental procedures are available in the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6495/1098/suppl/DC1
Materials and Methods
Figs. S1 to S6
HPLC Traces
NMR Spectra
References (*31*–*54*)

11 February 2020; accepted 17 April 2020
10.1126/science.abb2559

ANALYTICAL CHEMISTRY

Nontargeted mass-spectral detection of chloroperfluoropolyether carboxylates in New Jersey soils

John W. Washington[1]*, Charlita G. Rosal[1], James P. McCord[2], Mark J. Strynar[2], Andrew B. Lindstrom[2], Erica L. Bergman[3], Sandra M. Goodrow[4], Haile K. Tadesse[2], Andrew N. Pilant[2], Benjamin J. Washington[5], Mary J. Davis[1], Brittany G. Stuart[6], Thomas M. Jenkins[7]

The toxicity and environmental persistence of anthropogenic per- and poly-fluoroalkyl substances (PFAS) are of global concern. To address legacy PFAS concerns in the United States, industry developed numerous replacement PFAS that commonly are treated as confidential information. To investigate the distribution of PFAS in New Jersey, soils collected from across the state were subjected to nontargeted mass-spectral analyses. Ten chloroperfluoropolyether carboxylates were tentatively identified, with at least three congeners in all samples. Nine congeners are ≥$(CF_2)_7$. Distinct chemical formulas and structures, as well as geographic distribution, suggest airborne transport from an industrial source. Lighter congeners dispersed more widely than heavier congeners, with the most widely dispersed detected in an in-stock New Hampshire sample. Additional data were used to develop a legacy-PFAS fingerprint for historical PFAS sources in New Jersey.

Per- and poly-fluoroalkyl substances (PFAS) are anthropogenic compounds used to impart surfactant, antistaining, antisticking, and related properties to a wide array of consumer and industrial products. Spurred by concerns regarding potential toxicity and environmental persistence of long-chain PFAS (*1*–*5*), in 2006 the U.S. Environmental Protection Agency (EPA) and eight leading PFAS manufacturers and users negotiated a voluntary "PFOA Stewardship Program" in which the companies agreed to work toward the elimination of perfluorooctanoate acid (PFOA, or C8), as well as C8 precursors and related longer-chain homologs from emissions and product content by 2015. With establishment of the PFOA Stewardship Program, numerous PFAS manufacturers and users initiated efforts to develop substitute compounds for legacy long-chain PFAS, commonly settling on structures that are treated as confidential business information. With proliferation of these substitute PFAS, environmental chemists have set about attempting to identify them using nontargeted, high-resolution mass spectrometry (HRMS) to assemble formulas and likely structures from molecular-precursor and -fragment data (*6*). High mass-resolution enables chemists to identify those molecular formulas that have exact masses within a user-specified mass-error threshold, and molecular-fragment masses and spectra of the molecules help narrow possible formulas further, ideally informing molecular structure as well (*7*).

Among participants in the PFOA Stewardship Program, several have operated industrial facilities, ongoing or in the past, in or near densely populated New Jersey. As part of efforts to elucidate industrial chemical sources, chemical species, and distribution of legacy and possible substitute PFAS in New Jersey, in late 2017 the New Jersey Department of Environmental Protection (NJDEP) collected soil samples. For this survey, samples primarily were collected in southern New Jersey, where two PFOA Stewardship Program signatories are located: Solvay, in West Deptford Township, and DuPont (now Chemours), in Pennsville Township. Historically, Solvay produced polyvinylidene fluoride (PVDF), which entailed use of Surflon, a surfactant that contains C9, C11, and C13 (perfluorononanoate, perfluoroundecanoate, and perfluorotridecanoate) perfluorocarboxylates (PFCAs) (*8*). By contrast, the DuPont/Chemours facility manufactured and used fluorotelomers [compounds synthesized from perfluoroalkyl iodide, composed of perfluorinated-carbon straight chains such as $F(CF_2)_6$–, and usually two-hydrogen-bearing carbons, such as $–CH_2CH_2–$] from 1962 until no later than 2014 (*9*). Sampling transects were collected in the dominant downwind directions as recorded at nearby Philadelphia International Airport, and remote locations around the state were sampled as well (sampling campaign details are available in the supplementary materials). These samples were sent to the EPA, Office of Research and Development (ORD) laboratory in Athens, Georgia.

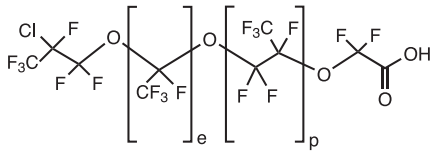


Fig. 1. A chloroperfluoropolyether carboxylate (ClPFPECA) identified by nontargeted MS analyses in soil samples from New Jersey. In the New Jersey samples, perfluoroethyl (e) plus perfluoropropyl (p) groups were observed to range in sum from one to four. The example congener depicted here would be designated (e,p) = 1,1. Isomers likely include an alternative terminal structure of $ClCF_2CF(CF_3)O–$ (*13*, *14*) as well as relative positions for the perfluoroethyl and perfluoropropyl groups.

At the ORD laboratory, soil samples were extracted (supplementary text, fig. S1, and table S1) in triplicate and selected samples analyzed (supplementary text) for PFAS unknown to our research team by using an ultra-performance liquid chromatograph (UPLC) coupled to a quadrupole time-of-flight (QToF) mass spectrometer operating in negative electrospray ionization (ESI), MS^e (no mass filtering) mode. Output data were sorted by signal intensity, high-intensity molecular features were plotted on mass-defect plots (*7*) ranging in defect from –0.10 to +0.05 Da, and molecular features appearing in the plots of multiple samples were selected for further scrutiny. Using low-collision-energy precursor masses, high-collision-energy fragment masses, a distinctive mono-chloro M+2 spectral feature, and carbon-isotopic ratios (*10*), we tentatively identified a molecular feature as a chloroperfluoropolyether carboxylate (ClPFPECA) that is described in the literature as "Solvay's product (CAS No. 329238-24-6)" (*11*), as reported in a product assessment by the European Food Safety Authority (EFSA) at the request of "Solvay Solexis, Italy" (*12*). With these reports, together with compound-synthesis papers by Solvay chemists (*13*, *14*), the structure of these ClPFPECAs appears to be as shown in Fig. 1 for 70% of production, with 30% having an alternative terminus of $ClCF_2CF(CF_3)O–$.

We have not had access to a standard of the Solvay product. However, on the basis of tentative identification of one Solvay product congener in our data, and the literature report that ClPFPECA congeners can include 0 to 2 perfluoroethyl groups (Fig. 1, e) and 1 to 4 perfluoropropyl groups (Fig. 1, p) (*11*, *12*) separated by ether linkages, we carried out suspect screening of our MS^e data by extracting hypothetical masses to determine what other congeners might be present. After

[1]U.S. Environmental Protection Agency (EPA), Office of Research and Development, Athens, GA, USA. [2]EPA, Office of Research and Development, Research Triangle Park, NC, USA. [3]New Jersey Department of Environmental Protection (NJDEP), Site Remediation and Waste Management Program, Trenton, NJ, USA. [4]NJDEP, Division of Science and Research, Trenton, NJ, USA. [5]EPA, Office of Research and Development, Washington, DC, USA. [6]EPA, Office of Research and Development, Cincinnati, OH, USA. [7]Senior Environmental Employment Program, Office of Research and Development, USEPA, Athens, GA, USA.
*Corresponding author. Email: washington.john@epa.gov

this effort, all tentatively identified congeners were further elucidated on the QToF operating in MS/MS mode, in which the quadrupole magnets were focused on suspected precursor mass/charge ratio (m/z) values and fragmented with ramped collision energy; then, precursors and fragments were isolated and detected in the ToF (supplementary text). Results for the nine ClPFPECA congeners tentatively identified on QToF are depicted in Fig. 2 and fig. S2. Within conventional HRMS-identification confidence context (*15*, *16*), these compounds fall at level 2b (diagnostic probable structure) and level 3 (tentative candidate), but considering the nine congeners together, confidence of their general identity is high.

Having tentatively identified nine congeners in these New Jersey soil samples as Solvay's product, we reexamined in-house nontargeted results for a water sample from the Bormida di Spigno River, downstream of Solvay Specialty Polymers Italy (Spinetta Marengo, Alessandria, Italy). In this Italian water sample, we identified five ClPFPECA congeners (fig. S3) that were consistent with our New Jersey soil samples, bolstering confidence still further in our identification of these compounds as Solvay's product.

Informed by the fragmentation patterns of the QToF suspect screening, we developed a method for routine analysis of the detected congeners on a conventional-resolution tandem mass spectrometer [liquid chromatography (LC)–MS/MS], adding monitoring for a possible ethyl,propyl (e,p)=1,0 congener (fig. S4 and table S2). Whereas this method was not developed with the benefit of authentic standards, it was informed by masses for ~30 precursors and fragments uniformly having mass error <4 mDa when the MS signal is $\geq 10^5$ (fig. S5). With an objective of assessing relative concentrations among samples, we performed analyses on the triplicate soil extracts with a matrix internal standard labeled with five heavy carbons, $^{13}C_5$-perfluorononanoic acid ($^{13}C_5$-PFNA; $^{13}C_5$-C9), then reported ClPFPECAs "as C9," by simple peak-area ratios (supplementary text). We also performed LC-MS/MS analyses on the triplicate soil-extract replicates for legacy PFCAs, quantitating on mass-labeled internal matrix standards (supplementary text). Results of ClPFPECA analyses are summarized in table S4, and PFCA analyses are summarized in table S5.

Of the 10 congeners we identified by means of QToF or tandem MS, (i) six were expected on the basis of EFSA information (e,p=0,1; 1,1; 0,2; 2,1; 1,2; and 0,3 congeners) (*11*, *12*); (ii) four were not included as congeners in the EFSA information (1,0; 2,0; 3,0; and 4,0 congeners); and (iii) six congeners anticipated on the basis of EFSA information were not detected (2,2; 1,3; 2,3; 0,4; 1,4; and 2,4 congeners) (fig. S6). In fig. S7, we summarize the fractional composition of the 10 ClPFPECA congeners detected in our study in terms of mean, maximum, and minimum fraction observed among our soil samples. Addressing the mean fractions, at roughly 40% each, the e,p = 0,1 and 1,1 congeners are dominant, followed by ~15% for the 0,2 and lesser to trace amounts of all other congeners (fig. S7).

Several ClPFPECAs eluted as split peaks (Fig. 2 and fig. S2). We investigated whether this splitting reflected the presence of isomers by extracting spectral patterns of visually distinct chromatographic peak ranges, looking for unique fragmentation patterns across aggregate peaks (supplementary text, qualitative examination for isomers, and figs. S8 to S10). On the basis of these efforts, we suspect the presence of group-regioisomerism for congeners having both ethyl and propyl groups as well as regioisomers based on chlorine position (Fig. 1).

These New Jersey soil samples generally were elevated in legacy PFCAs relative to global background soil estimates (*17*) and particularly

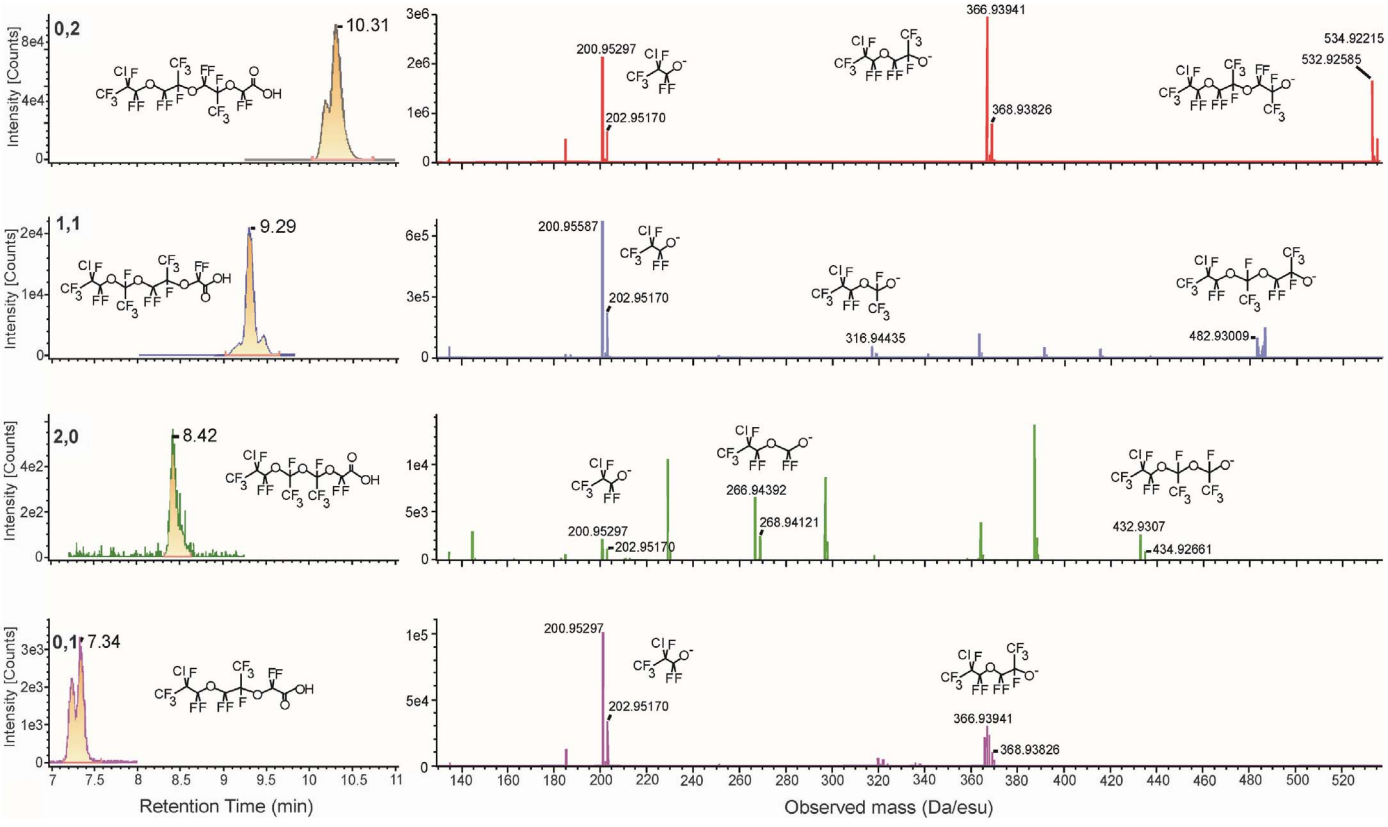


Fig. 2. Mass chromatograms (MS/MS mode), spectra, and precursor and fragment structures of four smaller ClPFPECA congeners detected in New Jersey samples. These are identified in the top left of the chromatograms by ethyl#,propyl#. Results for larger congeners are shown in fig. S2. Chromatogram peaks consist of signal from precursors and selected major fragments. Congeners elute in order according to molecular mass, small to large. On major spectra, the diagnostic monochlorine signal is 3:1 for ^{35}Cl:^{37}Cl.

elevated in C9 and longer homologs. For example, the mean C9 in our New Jersey soils is 785 pg/g dry soil (table S5) [compared with global background of 18 pg/g (*17*)]; mean C10 = 437 pg/g (perfluorodecanoate; background = 11 pg/g); mean C11 = 1618 pg/g (background = 9.6 pg/g); mean C12 = 167 pg/g (perfluorodecanoate; background = 9.0 pg/g); and mean C13 = 222 pg/g (background not reported). Also, the lowest New Jersey soil concentrations in our study for C9 through C12 PFCAs (table S5) were 5- to 30-fold that of mean global background values (*17*). These increased long-chain concentrations resulted in an anomalous PFCA-homolog profile for the New Jersey samples relative to global background. Whereas the PFCA profile for global background soils tended to be highest in C6, C7, and C8 PFCAs (perfluorohexanoate, perfluoroheptanoate, and perfluorooctanoate), in this order, these New Jersey samples were most highly represented by C11 and C9, in this order (fig. S11).

Taken altogether, these data for ClPFPECAs and the elevated levels of legacy PFAS strongly suggest the presence of regional PFAS sources.

Probing for possible relationships suggested by variation in the data, we performed principal component analysis (PCA) to guide directed testing (fig. S12). Principal component 1 (PC1) and PC2 account for 96.8% of variation in the data, with PC1 alone accounting for 90.6%. The 95% confidence interval ellipsoids in the PCA score plot (fig. S12) encompass the two chemical families almost exclusively: the ClPFPECAs and the legacy PFCAs. The major ellipsoidal axis of the ClPFPECA cluster is oriented more closely parallel to PC1, reflecting considerable variance among these data that can be characterized dominantly by a single component, as might be expected for a single physical source. Additionally, C11 and C13 fall within the ClPFPECA ellipsoid (fig. S12), suggesting similarities in the pattern of variation for C11 and C13 with at least some of the ClPFPECAs.

Exploring variation in the ClPFPECA data (fig. S12), we regressed the eight ClPFPECA congeners detected in most samples (excluding rarely detected 1,0 and 4,0 congeners) against distance from Solvay in log-transformed space (Fig. 3A). All eight congeners decreased with distance from Solvay with high degrees of significance ($P < 0.0002$) (Table 1). Examining the data in three dimensions, the ClPFPECA concentration contours form a concentric focus on Solvay, which is consistent with Solvay being the source of these compounds (Fig. 4). The slope of diminishing concentration with distance from Solvay (Table 1) also increases with molecular mass ($P < 0.001$) (Fig. 3B), suggesting that smaller congeners were dispersed more widely than larger congeners. This sorting by mass might be a factor in the absence of our detection of several of the largest ClPFPECA congeners expected for the Solvay product (fig. S6) (*12*); the heaviest congener we detected is the e,p = 0,3 at 792.9 Da, and the lightest of the six congeners expected, but not detected (fig. S6), was the 2,2, with a mass of 858.9 Da.

Considering that these soil samples chiefly are from positions that are not hydraulically downgradient in the watershed of any Solvay wastewater discharge (Fig. 4 and fig. S1), aqueous discharge cannot explain these observations, so these correlations strongly suggest atmospheric release from Solvay as the principal mode of occurrence for these soils.

The observation that three of the lightest congeners (0,1; 1,1; and 0,2) were detected in all study samples, including the most remote New Jersey sample near the northern state border (sample SS22) (fig. S1), suggests that light congeners might be dispersed beyond New Jersey state boundaries. To explore this possibility, we analyzed an in-stock sample from Merrimack, New Hampshire, that falls roughly parallel with the downwind transect extending northeasterly from Solvay (fig. S13). To determine whether unrelated samples might have ClPFPECAs, we also analyzed an in-stock

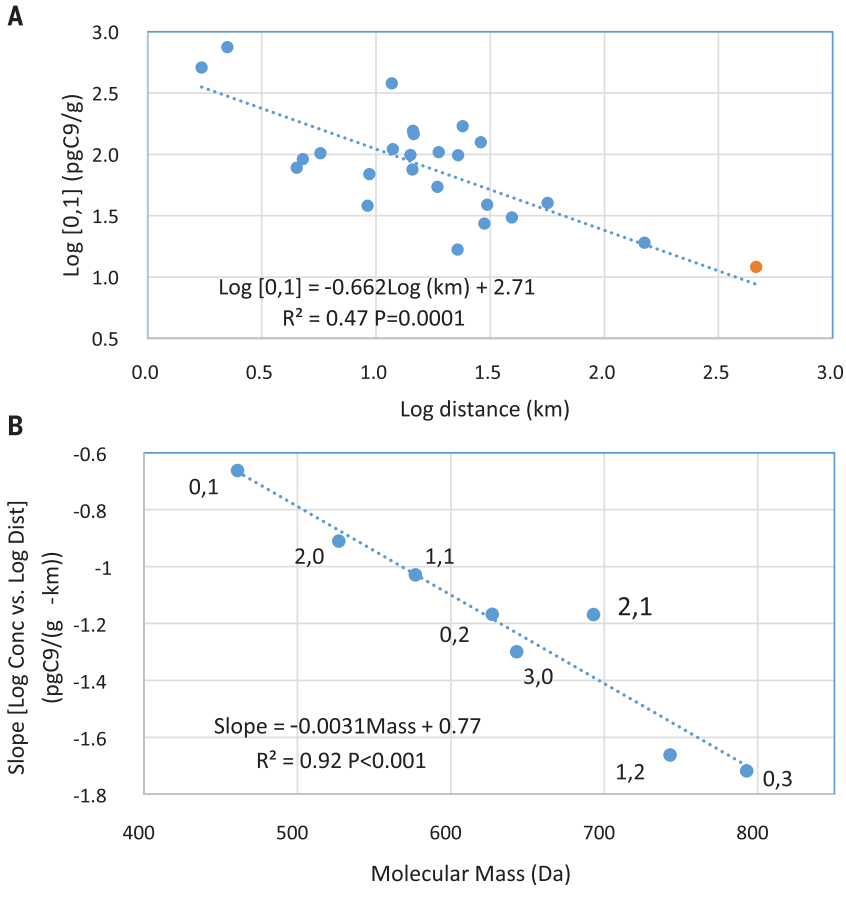


Fig. 3. Concentration profile. (**A**) Log 0,1-congener soil concentration (picograms/gram) versus log distance from Solvay (kilometers). The regression statistics are for the New Jersey soil samples (blue) located as far as 150 km removed from Solvay (table S1). Other ClPFPECA congeners are still more highly correlated with distance from Solvay (Table 1). Also shown is the 0,1 congener detected in a soil from Merrimack, New Hampshire, at 12.1 pg/g (orange), some 460 km distant from Solvay (table S1), falling closely proximate to the regression line for New Jersey 0,1 congeners. The 0,1 congener is the most widely dispersed of the ClPFPECAs (B) and the only ClPFPECA detected in the New Hampshire soil. Inclusion of the New Hampshire data point in the regression [coefficient of determination (R^2) = 0.55; $P = 10^{-5}$] increases the significance of the relationship roughly an order of magnitude beyond that of New Jersey data alone. (**B**) Regression slope (log [ClPFPECA] versus log distance from Solvay) for each of eight ClPFPECA congeners versus congener molecular mass. Given the statistically significant relationship ($P = 0.001$), this observation suggests sorting by molecular mass in an atmospheric plume, with lighter molecules generally being dispersed more remotely than heavier molecules. Mechanisms of atmospheric mass sorting remain uncertain, but the regression slope also is correlated with congener-acid vapor pressure ($R^2 = 0.91$; $P < 0.001$) and congener-anion octanol-water partition coefficient ($R^2 = 0.92$; $P < 0.001$), as estimated by the EPA Chemical Transformation Simulator (*25*).

Table 1. Regression statistics for chemical data (picograms/gram) against distance from selected facilities in log-transformed space. ND, not detected; *n*, sample count; Nonsig., nonsignificant; PFUA, perfluoroundecanoic acid; PFDA, perfluorodecanoic acid; PFDoA, perfluorododecanoic acid; PFTeA, perfluorotetradecanoic acid.

Analyte atmospheric precursor*	Distance from Solvay (km) (maximum *n* = 24)				Distance from Chemours (km) (anomalous background SS22 excluded; *n* = 23)			
	Compound(s)	Pearson *R*	*P*†	Slope	Compound(s)	Pearson *R*	*P*†	Slope
	0,1 (ND = 0, *n* = 24)	0.688	2.0×10^{-4}	−0.662				
	2,0 (ND = 2, *n* = 22)	0.766	3.2×10^{-5}	−0.911				
	1,1 (ND = 0, *n* = 24)	0.791	4.1×10^{-6}	−1.029				
	0,2 (ND = 0, *n* = 24)	0.845	2.0×10^{-7}	−1.167				
	3,0 (ND = 3, *n* = 21)	0.822	4.9×10^{-6}	−1.300				
	2,1 (ND = 2, *n* = 22)	0.831	1.7×10^{-6}	−1.169				
	1,2 (ND = 7, *n* = 17)	0.846	1.9×10^{-5}	−1.662				
	0,3 (ND = 4, *n* = 20)	0.849	2.2×10^{-6}	−1.718				
	ΣCongeners (ND = 0, *n* = 24)	0.796	3.3×10^{-6}	−0.937				
8:2FTOH	PFNA (C9)	0.130	Nonsig.		PFOA (C8)	0.202	Nonsig.	
10:2FTOH	PFUA (C11)	0.620	1.2×10^{-3}	−0.464	PFDA (C10)	0.514	1.2×10^{-2}	−0.404
12:2FTOH	PFTrA (C13)	0.482	1.7×10^{-2}	−0.356	PFDoA (C12)	0.478	2.1×10^{-2}	−0.394
14:2FTOH	(C15 not analyzed)				PFTeA (C14)	0.426	4.3×10^{-2}	−0.337
	(C9 + C11 + C13)	0.519	4.7×10^{-3}	−0.324	(C8 + C10 + C12)	0.204	Nonsig.	
	(C11 + C13)	0.604	1.8×10^{-3}	−0.449	(C10 + C12)	0.519	1.1×10^{-2}	−0.402
	(C9 + C11 + C13) − (C8 + C10 + C12)	0.383	Nonsig.					
	(C11 + C13) − (C10 + C12)	0.660	4.5×10^{-4}	−0.608				

*Source, Ellis *et al.* (*20*). †Significance level.

sample from Conyers, Georgia, which is roughly 1000 km southwest from Solvay (fig. S13). We detected the 0,1 congener in the downwind Merrimack sample (Fig. 3A) and no other congeners, and we detected no ClPFPECAs in the remote Conyers sample. The 0,1 congener is the most widely dispersed (Fig. 3B and Table 1), and the New Hampshire sample, some 450 km removed, plots closely proximate to the regression line for the 0,1 congener in New Jersey samples as a function of distance to Solvay (Fig. 3A). However, whether this New Hampshire 0,1-congener detection is from Solvay or some unknown source requires more study.

Given the role of Solvay as potentially the dominant or sole source of ClPFPECAs in our study, plots of legacy PFCAs against ClPFPECAs potentially guide which, if any, legacy PFCAs remain diagnostic of pre-Stewardship Solvay releases. Plotting concentrations of each legacy PFCA, chain lengths C4 (perfluorobutanoic acid) through C13 [perfluorotridecanoic acid (PFTrA)], against the sum of ClPFPECAs in fig. S14 shows three samples from closely proximate to Solvay that are high in ClPFPECAs also are high in C9, C11, and C13 PFCAs. On the basis of this observation, C9, C11, and C13 were regressed against distance from Solvay. Results of these regressions indicated that C9 is not correlated with distance from Solvay, but consistent with the PCA (fig. S12), C11 ($P = 1.2 \times 10^{-3}$) and C13 ($P = 1.7 \times 10^{-2}$) were statistically related with distance from Solvay (Table 1 and fig. S15). The seeming inconsistency of C9 plotting anomalously in fig. S14 but not being statistically related to distance from Solvay is likely due in large part to the relatively much higher mobility of C9 than C11 and C13 in soils. For example, in a study of PFCAs in Decatur, Alabama, soils, Washington *et al.* (*18*) reported deep-to-surface soil ratios for C9 as high as 50-fold that of C11 or C13, suggesting much higher rates of loss for C9 than C11 and C13 from surface soils through leaching and percolation.

Although figs. S14 and S15 and Table 1 suggest a relationship of C11 and C13 with Solvay, considerable spread remains in the data (fig. S15), perhaps reflecting noise imparted from other sources. The majority of all environmental releases of PFCAs longer than C8 from 1951 to 2015 arose from fluorotelomer- and C9-based products (*19*). According to smog-chamber experiments (*20*) and global-scale modeling that used a complex suite of kinetic constants estimated from literature (*21*), atmospheric oxidation of *n*:2FTOHs (where *n* is an even integer and FTOHs are fluorotelomer alcohols) yields roughly equimolar *n*PFCAs and (*n*+1)PFCAs or preferentially *n*PFCAs in urban areas where nitrogen oxides can be elevated. In soils, microbially mediated degradation of *n*:2FTOHs has been shown to proceed through beta oxidation to yield dominantly *n*PFCAs (*22*, *23*). Consistent with these studies, in their global soil survey, Rankin *et al.* (*24*) reported that C8/C9 [*n*PFCA/(*n*+1)PFCA] ratios commonly fall in roughly equimolar to dominantly C8 (*n*PFCA) range and argued atmospheric or soil degradation of fluorotelomers as a dominant mode of PFCAs occurrence globally. Given (i) historical production and use of fluorotelomers at the large-scale New Jersey Chemours facility, (ii) the generally prevalent contribution of fluorotelomers to C10 and C12, and (iii) atmospheric and soil fluorotelomer-degradation stoichiometry favoring roughly equimolar or dominantly even-chain PFCAs, the difference of *n*PFCAs minus (*n*+1)PFCAs, (C11 + C13) − (C10 + C12), has the potential to deconvolute potential signals from Solvay and Chemours for these legacy PFAS. Large positive excesses in this difference suggest direct release of C11 and C13 PFCAs, whereas near-zero or negative values of this difference would be consistent with atmospheric or soil degradation of fluorotelomer precursors as a source.

Applying the difference (C11 + C13) − (C10 + C12) to our New Jersey soil data accentuates signal to noise in that the strength of correlation with distance from Solvay (fig. S16) increases nearly an order of magnitude beyond that of C11 or C13 alone, with $P = 4.5 \times 10^{-4}$ (Table 1). (C11+C13) − (C10+C12) is plotted in fig. S17 as a function of the sum of ClPFPECAs, illustrating a relationship significant at $P = 4.0 \times 10^{-5}$ and bolstering that these parameters reflect a common mode of occurrence: airborne transport.

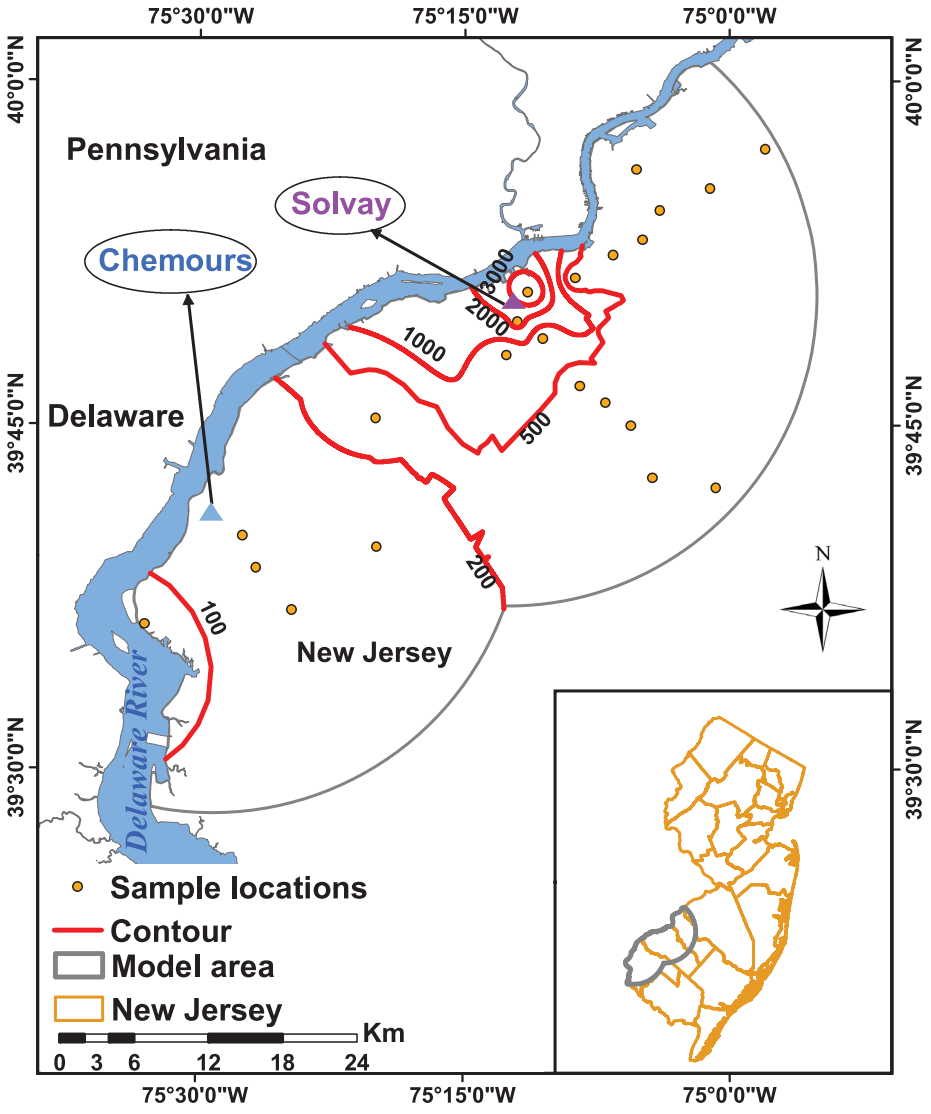


Fig. 4. Geographic distribution. Shown are ΣClPFPECAs in surface soils (picograms/gram). Contour lines were generated by using an algorithm in ArcMAP 10.6.1 that weighted the five nearest data points according to inverse-square distance. Despite some geographic sporadicity in the data and numerical artifacts where data are sparsely spaced, taken as a group the contours depict a clear pattern of increasing ΣClPFPECAs with proximity to Solvay.

Contours of the difference (C11 + C13) – (C10 + C12) are mapped in fig. S18. The resulting pattern depicts a strongly expressed positive anomaly focusing on Solvay as well as a negative anomaly proximate to Chemours, a pattern that is consistent with the reasoning above (fig. S18). These results are consistent with values reported in Rankin *et al.* (*24*) in that three of four samples collected ~20 km southeast of Chemours calculate to negative values for the difference (C11 + C13) – (C10 + C12). Taken altogether then, the difference (C11 + C13) – (C10 + C12) evidently fingerprints two potential PFAS sources in concert by accentuating differences in mode of occurrence: direct odd-chain PFCA release from the Solvay facility versus fluorotelomer degradation in the atmosphere or soil from the Chemours facility.

Here, we have reported tentative identification of 10 ClPFPECA congeners distributed across an expansive breadth of soils in densely populated New Jersey and likely beyond. In light of these findings, numerous near-term pressing uncertainties merit investigation, including the presence and mobility of the congeners in soil profiles, in surface and ground waters, in vegetation (such as agricultural crops), and in animals including humans, as well as whether there is evidence that these ClPFPECAs degrade in the environment. In the longer term, investigation of whether these ClPFPECAs might be toxic is prudent.

REFERENCES AND NOTES

1. J. P. Giesy, K. Kannan, *Environ. Sci. Technol.* **35**, 1339–1342 (2001).
2. A. O. De Silva, S. A. Mabury, *Environ. Sci. Technol.* **38**, 6538–6545 (2004).
3. C. Lau, J. L. Butenhoff, J. M. Rogers, *Toxicol. Appl. Pharmacol.* **198**, 231–241 (2004).
4. K. Kannan *et al.*, *Arch. Environ. Contam. Toxicol.* **48**, 559–566 (2005).
5. M. Houde, J. W. Martin, R. J. Letcher, K. R. Solomon, D. C. G. Muir, *Environ. Sci. Technol.* **40**, 3463–3473 (2006).
6. X. Lim, *Nature* **566**, 26–29 (2019).
7. S. Newton *et al.*, *Environ. Sci. Technol.* **51**, 1544–1552 (2017).
8. K. Prevedouros, I. T. Cousins, R. C. Buck, S. H. Korzeniowski, *Environ. Sci. Technol.* **40**, 32–44 (2006).
9. AECOM, "Conceptual site model (CSM) for poly- and perfluoroalkyl substances (PFAS) prepared by AECOM on behalf of The Chemours Company" (AECOM, 2017).
10. J. W. Washington, C. G. Rosal, E. M. Ulrich, T. M. Jenkins, *J. Chromatogr. A* **1583**, 73–79 (2019).
11. Z. Wang, I. T. Cousins, M. Scheringer, K. Hungerbühler, *Environ. Int.* **60**, 242–248 (2013).
12. EFSA, *EFSA J.* **8**, 1519 (2010).
13. C. Tonelli, A. Di Meo, R. Picozzi, *J. Fluor. Chem.* **128**, 46–51 (2007).
14. C. Tonelli, A. Di Meo, R. Picozzi, M. Bassi, *J. Fluor. Chem.* **132**, 356–362 (2011).
15. L. W. Sumner *et al.*, *Metabolomics* **3**, 211–221 (2007).
16. E. L. Schymanski *et al.*, *Environ. Sci. Technol.* **48**, 2097–2098 (2014).
17. J. W. Washington, K. Rankin, E. L. Libelo, D. G. Lynch, M. Cyterski, *Sci. Total Environ.* **651**, 2444–2449 (2019).
18. J. W. Washington, H. Yoo, J. J. Ellington, T. M. Jenkins, E. L. Libelo, *Environ. Sci. Technol.* **44**, 8390–8396 (2010).
19. Z. Wang, I. T. Cousins, M. Scheringer, R. C. Buck, K. Hungerbühler, *Environ. Int.* **70**, 62–75 (2014).
20. D. A. Ellis *et al.*, *Environ. Sci. Technol.* **38**, 3316–3321 (2004).
21. T. J. Wallington *et al.*, *Environ. Sci. Technol.* **40**, 924–930 (2006).
22. N. Wang *et al.*, *Chemosphere* **75**, 1089–1096 (2009).
23. J. W. Washington, T. M. Jenkins, E. J. Weber, *Environ. Sci. Technol.* **49**, 13256–13263 (2015).
24. K. Rankin, S. A. Mabury, T. M. Jenkins, J. W. Washington, *Chemosphere* **161**, 333–341 (2016).
25. U.S. EPA, *CTS: On-line Chemical Transformation Simulator* (EPA, 2019).
26. U.S. EPA, New Jersey NJDEP/EPA-ORD soil PFAS study. EPA Environmental Dataset Gateway (2019); doi:10.23719/1506012.

ACKNOWLEDGMENTS

The authors thank N. Azzam, T. Buckley, T. Collette, A. Gillespie, M. Medina-Vera, G. Post, B. Schumacher, C. Stevens, K. Sullivan, and E. Weber for support and technical review. **Funding:** This research was supported by the EPA ORD and the NJDEP, Site Remediation and Waste Management Program. It has been subjected to the EPA's administrative review and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use. The views expressed in this article are those of the authors and do not necessarily represent the views or policies of the EPA. J.W.W. is an adjunct faculty member at the University of Georgia, Department of Geology. **Author contributions:** J.W.W. analyzed samples, detected and identified the ClPFPECAs in soil, quantitated analytes, interpreted results, and drafted the manuscript. C.G.R. identified ClPFPECAs. J.P.M. and M.J.S. independently detected and identified ClPFPECAs in water samples from New Jersey and Italy. A.B.L. organized and led technical design of the study for the EPA. E.L.B. and S.M.G. organized and designed the study for NJDEP and oversaw the sampling campaign. H.K.T. and A.N.P. performed spatial contouring and developed maps. B.J.W. carried out and led interpretation of the PCA. M.J.D. integrated data for the ClPFPECAs. B.G.S. reviewed data quality and integrity. T.M.J. extracted the samples. **Competing interests:** The authors declare no competing interests nor outside consulting. **Data and materials availability:** All data included in this study are available in the main text and in the supplementary materials; underlying data are available on-line at the EPA Environmental Dataset Gateway (*26*) (DOI: 10.23719/1506012).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6495/1103/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S18
Tables S1 to S5
References (*27*, *28*)

31 December 2019; accepted 16 April 2020
10.1126/science.aba7127

NEUROSCIENCE

Restoring light sensitivity using tunable near-infrared sensors

Dasha Nelidova[1,2], Rei K. Morikawa[1,2], Cameron S. Cowan[1,2], Zoltan Raics[1,2], David Goldblum[3], Hendrik P. N. Scholl[1,3,4], Tamas Szikra[1,2], Arnold Szabo[5], Daniel Hillier[1,2,6,7,8]*, Botond Roska[1,2,3]*

Enabling near-infrared light sensitivity in a blind human retina may supplement or restore visual function in patients with regional retinal degeneration. We induced near-infrared light sensitivity using gold nanorods bound to temperature-sensitive engineered transient receptor potential (TRP) channels. We expressed mammalian or snake TRP channels in light-insensitive retinal cones in a mouse model of retinal degeneration. Near-infrared stimulation increased activity in cones, ganglion cell layer neurons, and cortical neurons, and enabled mice to perform a learned light-driven behavior. We tuned responses to different wavelengths, by using nanorods of different lengths, and to different radiant powers, by using engineered channels with different temperature thresholds. We targeted TRP channels to human retinas, which allowed the postmortem activation of different cell types by near-infrared light.

Photoreceptor degeneration, including age-related macular degeneration and retinitis pigmentosa, is the leading cause of blindness in industrialized countries. When cone photoreceptors lose light sensitivity, high-resolution vision is affected, and it becomes difficult to carry out the activities of daily living. In most cases, photoreceptor degeneration is incomplete, leading to the presence of light-sensitive and light-insensitive photoreceptor zones next to each other within the same retina. Remaining light-sensitive regions limit the utility of optogenetic (*1*) or light-switch (*2*) therapies, because these technologies require bright, visible light that saturates photoreceptors.

Enabling the detection of near-infrared (NIR) light (>900 nm) at wavelengths outside the spectrum visible to the human eye (390 to 700 nm) could provide a way of supplementing or restoring light sensitivity in the affected retinal region without interfering with the vision that remains. Currently, there is no technology that would allow the induction of NIR sensitivity in a blind retina.

A few species, such as boas, pythons, and pit vipers, can detect infrared light (1 to 30 μm) using temperature-sensitive transient receptor potential (TRP) cation channels expressed in a specialized organ (*3*). Thermal and visual images superimpose in the snake's brain (*4*), presumably enabling the snake to react to the environment with greater precision than what is possible using only a single image. TRP channels could potentially be targeted to retinal cell types to make them sensitive to infrared radiation. However, heat transfer to ectopically expressed TRP channels via direct NIR illumination is inefficient, requiring high intensities that would damage the retina.

To develop a more efficient NIR light detector for retinal cell types, we engineered a dual system consisting of a genetic component and a nanomaterial component (Fig. 1A). The genetic component consisted of temperature-sensitive TRP channels, engineered to incorporate an extracellular epitope recognizable by a specific antibody (*5*). The nanomaterial component consisted of gold nanorods conjugated to an antibody against the epitope (*6*). This system uses surface plasmon resonance for heat transfer (*7*): Gold nanorods capture NIR light at their resonant wavelength and produce heat, which is harnessed to open TRP channels in the proximity of the nanorods. The epitope ensures nanorod binding to engineered rather than native TRP channels, because some TRP channels are expressed in the retina (*8*, *9*).

We developed a system based on rat TRP family V member 1 (TRPV1) channels and gold nanorods with absorption maxima (λ_{abs}) at 915 nm; this value was selected to ensure low water absorption. We inserted a 6x-His epitope tag in the middle of the first TRPV1 extracellular loop (Fig. 1, C and D), after amino acid 459 or 465 (fig. S1). Analysis of TRPV1 structure suggested that insertion at these sites would not disrupt protein function.

To measure whether tagged channels were functional, we performed whole-cell voltage clamp in human embryonic kidney (HEK) cells expressing TRPV1.459-6x-His, TRPV1.465-6x-His, or untagged TRPV1 while activating the channels by TRPV1 agonist capsaicin. The sizes of evoked currents were similar between TRPV1.465-6x-His and TRPV1 (table S1, row A) but smaller in TRPV1.459-6x-His (table S1, row B, and fig. S2). Therefore, we used TRPV1.465-6x-His (abbreviated as rTRPV1) in subsequent experiments.

We targeted rTRPV1 to cone photoreceptors of *Pde6b*^{rd1} mice (known as rd1 mice) through subretinal injection of adeno-associated virus (AAV), using a photoreceptor-specific mouse cone arrestin promoter (mCar) to restrict expression (Fig. 1E). Rd1 mice have severe photoreceptor degeneration, with complete loss of rods and dysfunctional, light-insensitive cone photoreceptors by 4 weeks of age (Fig. 1B) (*10*). Cell membrane expression of rTRPV1 was seen in 55 ± 10% of rd1 cones (table S1, row C, and Fig. 1F). Among rTRPV1-positive cells, 98 ± 1.6% were cones (table S1, row D); these rTRPV1-positive cones expressed the 6x-His tag (table S1, row E, and Fig. 1, E and F).

To measure whether NIR light drives responses in rd1 retinas, we performed two-photon calcium imaging of individual cone cell bodies and axon terminals as well as ganglion cell bodies in wholemount P56-P72 retinas under two conditions: (i) rTRPV1 with nanorods (λ_{abs} = 915 nm) and (ii) rTRPV1 without nanorods. To measure whether NIR light affects normal cones, we performed two-photon calcium imaging of cone axon terminals in wholemount wild-type retinas stimulating cones with NIR (915 nm) and/or visible light. To detect calcium signals in cones, we genetically targeted the calcium indicator GCaMP6s via an AAV that expresses GCaMP6s under a cone-specific promoter (*11*). For ganglion cells, we used the organic calcium sensor Oregon Green 488 BAPTA-1 (OGB-1).

rTRPV1-expressing rd1 cones showed 915-nm light ("NIR light")–evoked increases of calcium signal in the presence of nanorods ("NIR cone response") (Fig. 1, G to I). NIR cone response was of opposite polarity compared with the visible light response of wild-type cones (Fig. 1G). Polarity reversal was due to cation selectivity of rTRPV1. NIR cone response was similar in size to the visible light response of wild-type cones (table S1, rows F and G). NIR light neither activated wild-type cones nor affected their visible light responses (fig. S3). rTRPV1-expressing cones without nanorods did not react to light (Fig. 1, G to I). In the presence of nanorods, NIR light also induced responses in neurons of the ganglion cell layer (GCL) (fig. S4). In all subsequent experiments, we used both the TRP channel and the nanorod component ("NIR sensor"); uninjected rd1 mice were used as the control.

To assess whether NIR light–induced retinal activity propagates to higher visual centers, we generated rd1 mice with targeted GCaMP6s expression in layer 4 of the primary visual cortex (V1). Layer 4 receives feedforward connections from the lateral geniculate nucleus. We performed two-photon calcium imaging

[1]Institute of Molecular and Clinical Ophthalmology Basel, Basel, Switzerland. [2]Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland. [3]Department of Ophthalmology, University of Basel, Basel, Switzerland. [4]Wilmer Eye Institute, Johns Hopkins University, Baltimore, MD, USA. [5]Department of Anatomy, Semmelweis University, Budapest, Hungary. [6]Deutsches Primatzentrum, Leibniz Institute for Primate Research, Göttingen, Germany. [7]Research Centre for Natural Sciences, Budapest, Hungary. [8]Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Budapest, Hungary.
*Corresponding author. Email: daniel.hillier@iob.ch (D.H.); botond.roska@iob.ch (B.R.)

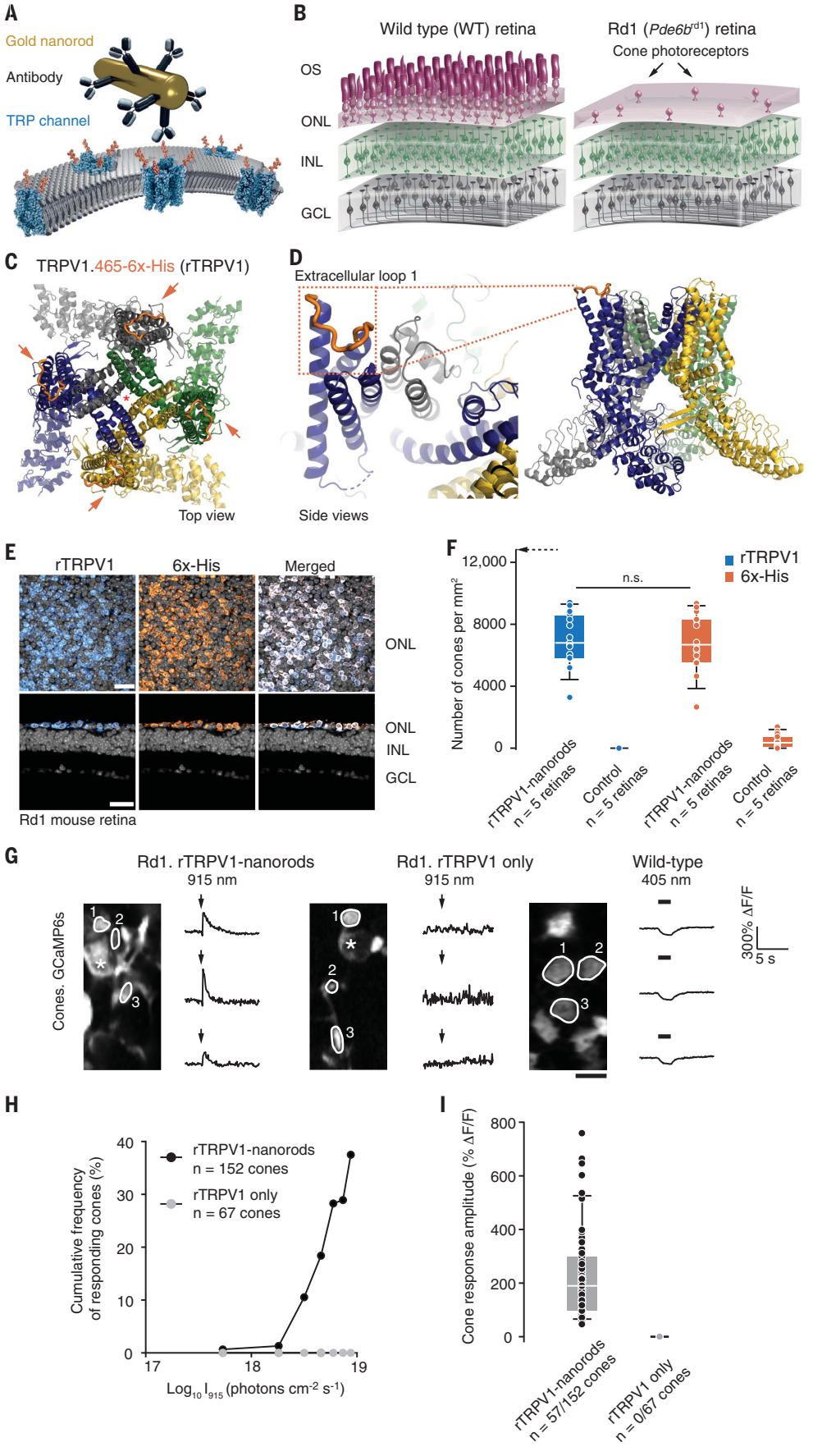


Fig. 1. NIR light responses in mouse cone photoreceptors. (**A**) Components of the NIR light sensor. Engineered TRP channels (blue) express protein epitope tags (orange) in extracellular domains and bind antibody-conjugated gold nanorods. (**B**) Left: Healthy retina. Photons are captured by outer segments (OS) of photoreceptor cells. Right: retinal degeneration, characterized by loss of OS and blindness. In the rd1 mouse model of degeneration, rod cell bodies are lost but cone cell bodies persist. ONL, outer nuclear layer; INL, inner nuclear layer; GCL, ganglion cell layer. (**C** and **D**) Structure of rTRPV1 channel. (C) Top view. Orange arrows indicate a 6x-His epitope tag in each of the four subunits of the TRPV1 tetramer (yellow, blue, gray, and green). The red asterisk marks the channel pore. (D) Right: Side view. Left: 6x-His epitope tag (orange) within the first extracellular loop of the blue subunit is enlarged. (**E**) Top row: Top views of rd1 retinas transduced with both rTRPV1 and nanorods and immunostained for TRPV1 (left, blue), 6x-His (middle, orange), and merging the two (right). Gray, Hoechst nuclear stain. Bottom row: Cross sections of the retinas shown in the top row. Scale bars, 25 μm. (**F**) Number of rTRPV1-positive (blue) and 6x-His–positive (orange) cones per square millimeter in rd1 retinas transduced with both rTRPV1 and nanorods (n = 5 mice) or in control, uninjected rd1 retinas (n = 5 mice). Dashed arrow indicates maximum cone density in rd1 mice at postnatal day 70 (*10*). Each data point is collected from a different region of a retina (three regions per retina). (**G**) Example calcium responses (mean $\Delta F/F$, where F represents baseline fluorescence; two to three repetitions) recorded from cone axon terminals in P56-P71 rd1 mice transduced with rTRPV1 and nanorods (left, λ_{abs} nanorod = 915 nm), in rd1 mice transduced with rTRPV1 only (middle), or in wild-type mice (right). TRP, n = 4 mice; wild-type, n = 3 mice. Stimulus, full-field NIR light (915 nm, $\log_{10}$ light intensity = 18.9; left and middle) or visible light (405 nm, $\log_{10}$ light intensity = 14; right). Black bars (2 s) and arrows (100 ms) indicate stimulus timing. Two-photon images of GCaMP6s-expressing cone axon terminals (white circles) are shown to the left of the response curves. Scale bar, 5 μm. White asterisks indicate cell bodies. (**H**) Cumulative frequency of responding rTRPV1-transduced rd1 cones with (black) and without (gray) nanorods (λ_{abs} = 915 nm). (**I**) Cone response amplitudes ($\Delta F/F$). Light intensities as in (G). Number of cones given as a ratio of responding cones to measured cones.

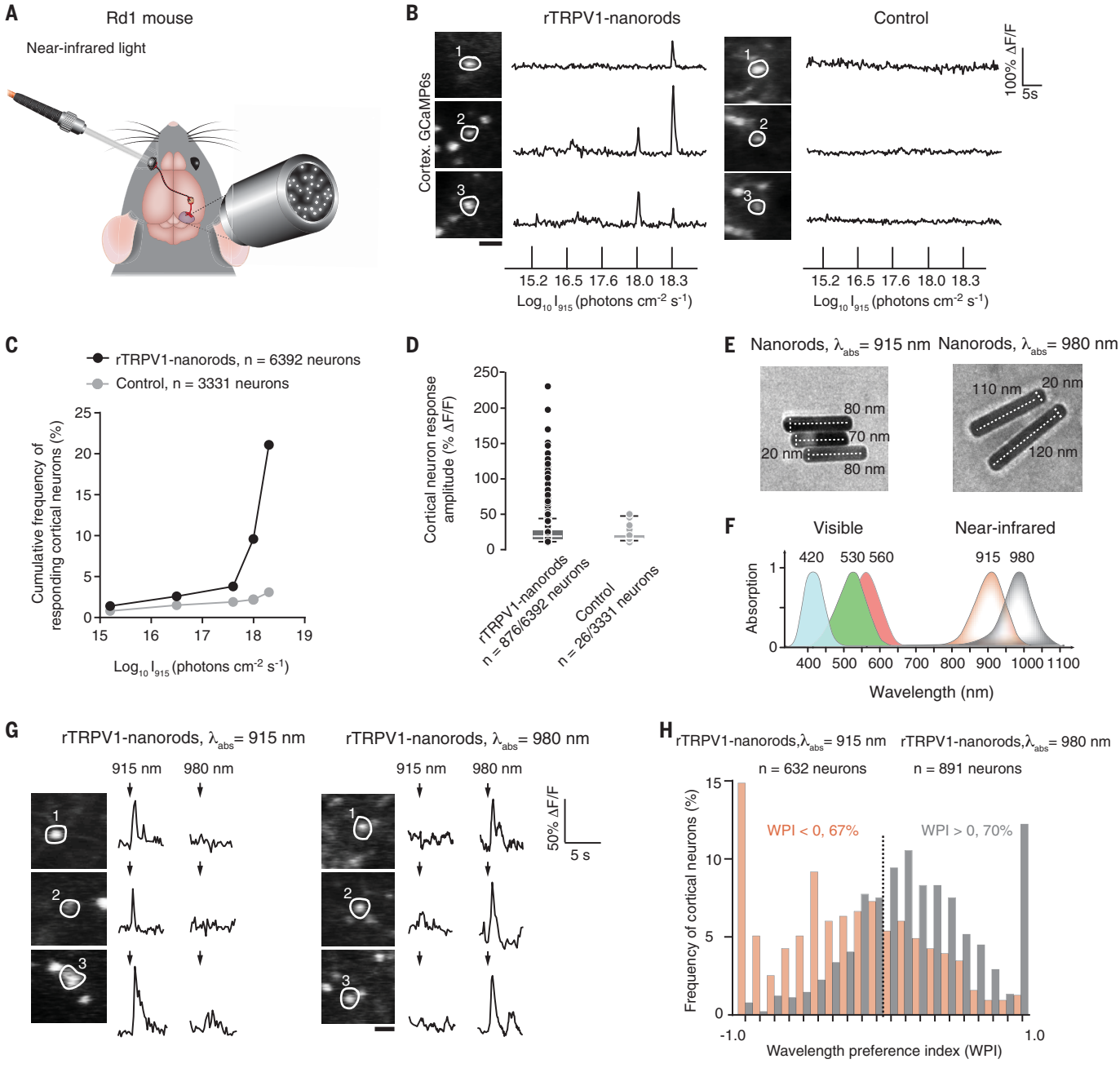


Fig. 2. NIR light responses in mouse primary visual cortex. (**A**) Schematic of the experiments in the primary visual cortex (V1). Cortical neuron calcium responses to 100-ms full-field NIR light stimulation of the contralateral eye were recorded in P51-P83 rd1 mice. (**B**) Example calcium responses (mean $\Delta F/F$, five repetitions) to 915-nm light stimulation recorded in rd1 mice transduced with rTRPV1 and nanorods (left, λ_{abs} = 915 nm, n = 3 mice) and in control, uninjected rd1 mice (right, n = 5 mice). Vertical lines at bottom indicate stimulus timing. Two-photon images of GCaMP6s-expressing neuronal cell bodies in layer 4 of V1 (white circles) are shown to the left of the response curves. Scale bar, 25 μm. (**C**) Cumulative frequency of responding cortical neurons in mice transduced with both rTRPV1 and nanorods (black, λ_{abs} = 915 nm) and in control, uninjected mice (gray). (**D**) Cortical neuron response amplitudes ($\Delta F/F$). Log_{10} light intensity = 18.3. Number of neurons given as ratio of responding neurons to measured neurons. (**E**) Morphology of nanorods tuned to 915 nm (left) and 980 nm (right), measured by transmission electron microscopy. (**F**) Schematic showing nanorod absorption spectra relative to visual pigment of the human retina. (**G**) Example cortical calcium responses (mean $\Delta F/F$, five repetitions) to 915-nm and 980-nm light stimulation (log_{10} light intensity = 18.3). Left: Mice transduced with rTRPV1 and nanorods with λ_{abs} = 915 nm (n = 3 mice). Right: Mice transduced with rTRPV1 and nanorods with λ_{abs} = 980 nm (n = 4 mice). Arrows indicate stimulus timing. Two-photon images of GCaMP6s-expressing neuronal cell bodies in layer 4 of V1 (white circles) are shown to the left of the response curves. Scale bar, 25 μm. (**H**) Frequency of cortical neurons as a function of the WPI in mice transduced with rTRPV1 and nanorods with λ_{abs} = 915 nm (orange) or nanorods with λ_{abs} = 980 nm (gray). For nanorods with λ_{abs} = 915 nm, larger fraction of 915 nm (WPI < 0) over 980 nm (WPI > 0) light preferring neurons. For nanorods with λ_{abs} = 980 nm, larger fraction of 980 nm (WPI > 0) over 915 nm (WPI < 0) light preferring neurons. Light intensities as in (G).

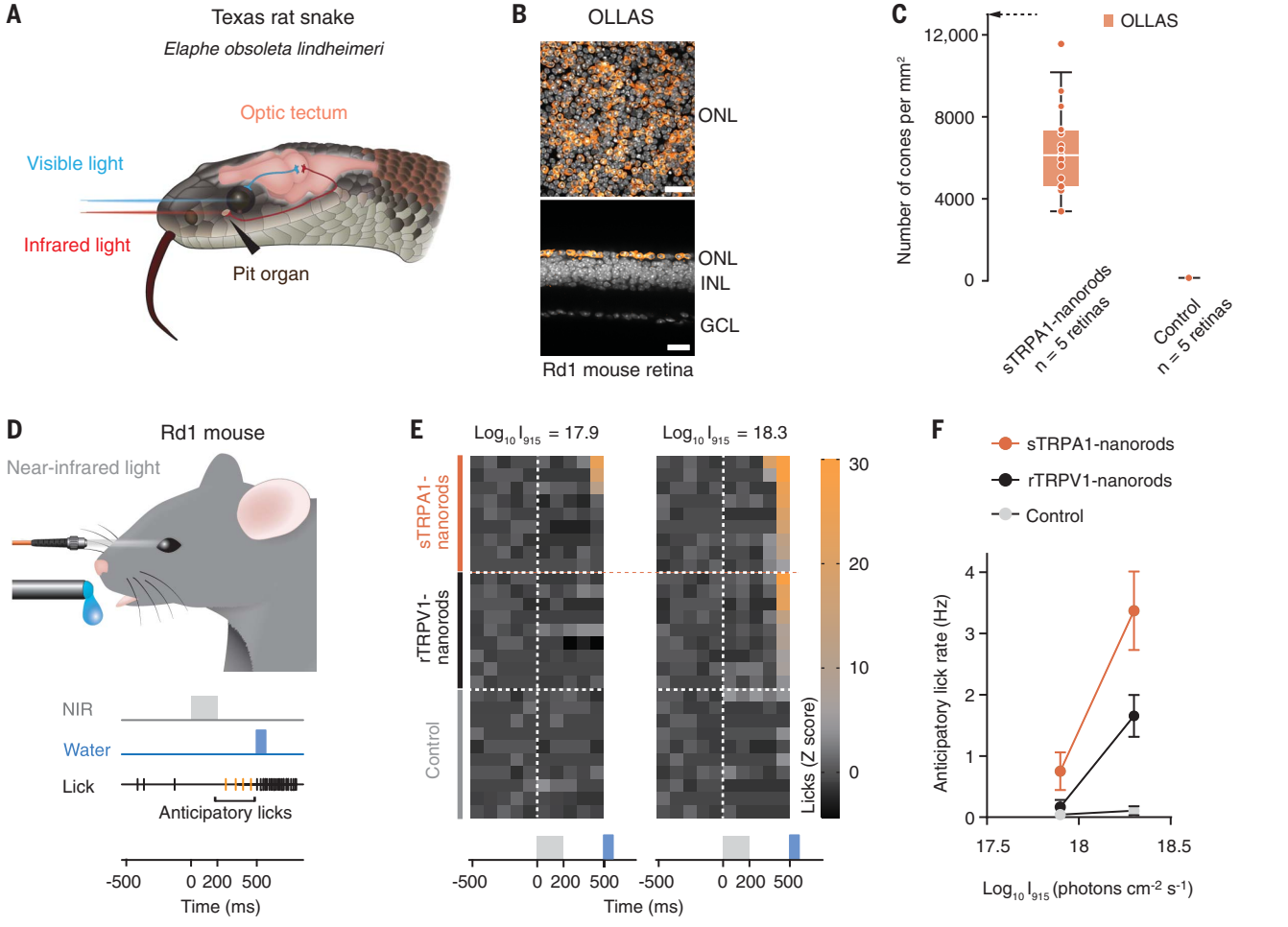


Fig. 3. NIR light–guided mouse behavior. (**A**) Schematic showing location of TRP channel–expressing, infrared-sensitive pit organ. Information is overlaid in the optic tectum. (**B**) Expression of sTRPA1 in cones of rd1 mice. Top: Top view of a retina transduced with both sTRPA1 and nanorods, immunostained for OLLAS (orange) and overlaid with Hoechst nuclear stain (gray). Bottom: Cross section of the retina shown in top image. Scale bars, 25 μm. (**C**) Number of OLLAS-positive (orange) cones per square millimeter in rd1 retinas transduced with both sTRPA1 and nanorods (n = 5 mice) or in control, uninjected rd1 retinas (n = 5 mice). Dashed arrow indicates maximum cone density in rd1 mice at postnatal day 70 (*10*). Each data point is collected from a different region of a retina (three regions per retina). (**D**) Schematic of behavioral task. NIR full-field stimulation of one eye (915 nm, 200 ms) cues water presentation for head-fixed, water-restricted P56-P73 rd1 animals. Mice respond by licking before (anticipation) or after the appearance of water. (**E**) Lick response heatmaps. Rows: Responses of different mice. Columns: Responses in 100-ms time bins. Top: rd1 mice transduced with both sTRPA1 and nanorods (λ_{abs} = 915 nm, n = 9 mice). Middle: rd1 mice transduced with both rTRPV1 and nanorods (λ_{abs} = 915 nm, n = 9 mice). Bottom: Control, uninjected rd1 mice (n = 10 mice). Left: Stimulus $\log_{10}$ light intensity = 17.9. Right: Stimulus $\log_{10}$ light intensity = 18.3. (**F**) Mean anticipatory lick rates quantified from (E) as a function of light intensity. Error bars, SEM.

in vivo in P51-P83 animals, recording layer 4 activity at single-cell resolution during NIR light stimulation of the eye (Fig. 2A). In NIR sensor–injected animals, cortical neurons showed NIR light–evoked increases in calcium signal (Fig. 2B), which were light-intensity dependent (Fig. 2C and fig. S5). Neuronal activation was greater in NIR sensor–injected animals than in controls (table S1, row H, and Fig. 2C).

Nanorod absorption spectra can be wavelength-tuned by varying nanorod aspect ratios (length-to-width ratios) (Fig. 2E and fig. S6). To test whether the action spectra of neuronal activity can also be tuned, we selected a second type of gold nanorod with peak absorption at 980 nm (aspect ratio: 5.5) and compared with nanorods with peak absorption at 915 nm (aspect ratio: 4.0). Both types were paired to rTRPV1. For each nanorod type, we performed layer 4 cortical calcium imaging in P51-P71 rd1 mice twice: once with 980-nm stimulation of the eye, and once with 915-nm stimulation. To classify cortical neurons as 980 nm or 915 nm responsive, we computed a wavelength preference index (WPI) for each NIR light–responsive neuron. We found a preference for 980-nm light over 915-nm light using nanorods tuned to 980 nm (table S1, row I; Fig. 2H; and fig. S7). Similarly, in animals with nanorods tuned to 915 nm, more cortical neurons preferred 915-nm light over 980-nm light (table S1, row J; Fig. 2H; and fig. S7).

Next, we asked whether molecular components can be tuned to increase sensitivity. A variety of TRPA1 channels also serve as heat sensors. TRPA1 from the Texas rat snake (*Elaphe obsoleta lindheimeri*) is activated at a lower temperature than rTRPV1 (*3*). To determine the suitability of snake TRPA1 as a NIR sensor component, we first engineered TRPA1 to express the peptide epitope tag OLLAS (*12*) within the first or second extracellular loop. Anti-OLLAS antibodies show improved immunodetection compared with anti-6x-His and other antibodies for conventional epitope tags (*12*). To identify loop domains, we first determined the location of extracellular loop domains of human TRPA1 from its cryo–electron microscopy structure (*13*). Subsequently, we identified potential loop domains of snake TRPA1 after pairwise sequence

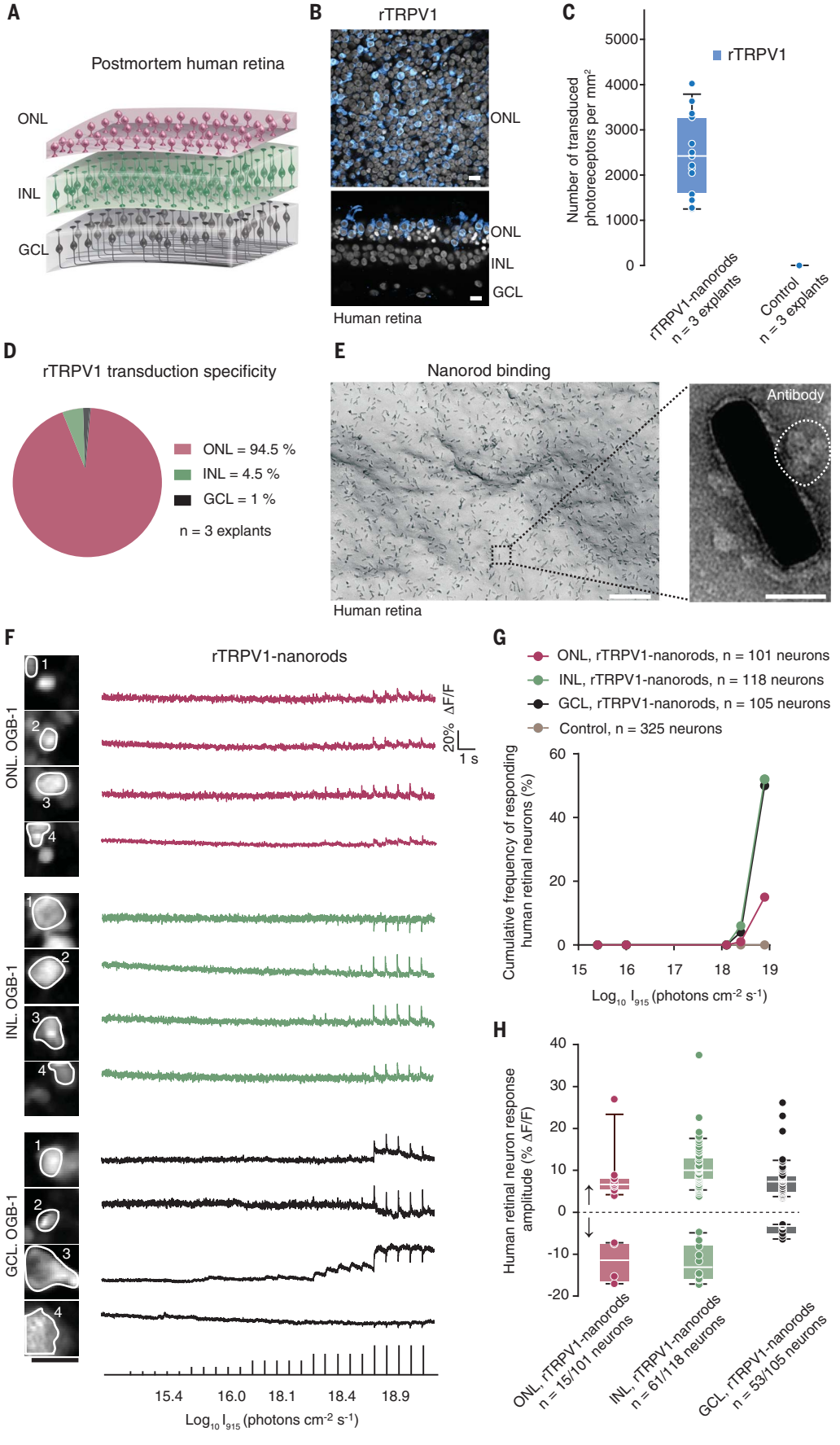


Fig. 4. NIR light responses in the ex vivo human retina. (**A**) Schematic of a human retinal explant 8 weeks postmortem. Long-term culture leads to loss of outer segments and light responses. (**B**) Top: Top view of a human retina transduced with both rTRPV1 and nanorods, immunostained for TRPV1 (blue) and overlaid with Hoechst nuclear stain (gray). Bottom: Cross section of the retina shown in top image. Scale bars, 10 μm. (**C**) Number of rTRPV1-positive (blue) photoreceptors per square millimeter in human retinas transduced with both rTRPV1 and nanorods or in control, untransduced human retinas. Each data point is collected from a different region of a retina (five regions per retina). (**D**) Distribution of rTRPV1-positive cells across retinal layers. (**E**) Scanning electron microscopy image of gold nanorods bound to a rTRPV1-transduced human retina. Scale bar, 1 μm. Inset: Transmission electron microscopy image showing an anti-6x-His antibody touching a gold nanorod with its Fc domain. Scale bar, 20 nm. (**F**) Example calcium responses recorded from human retinal neurons in ONL, INL, and GCL in response to full-field NIR light. Vertical lines at bottom indicate stimulus (915 nm, 100 ms) timing. Two-photon images of OGB-1 filled cell bodies (white circles), left of the response curves. Scale bar, 10 μm. (**G**) Cumulative frequency of responding neurons in different layers of human retinas transduced with both rTRPV1 and nanorods (λ_{abs} = 915 nm) or in control, untransduced human retinas. (**H**) Response amplitudes ($\Delta F/F$) by retinal layer. Increase (↑) and decrease (↓) of calcium signal are shown separately. Number of neurons given as ratio of responding neurons to measured neurons. Log_{10} light intensity = 18.9.

alignment (*14*) between human and snake sequences. OLLAS was placed after amino acid 755 or 758, corresponding to the first loop, or after amino acid 824, corresponding to the second loop (fig. S8).

We performed whole-cell voltage clamp in HEK cells expressing TRPA1.755-OLLAS, TRPA1.758-OLLAS, TRPA1.824-OLLAS, or untagged TRPA1 while activating the channels by TRPA1 agonist allyl isothiocyanate. The sizes of evoked currents were similar between TRPA1.755-OLLAS and TRPA1 (table S1, row K) and between TRPA1.824-OLLAS and TRPA1 (table S1, row L, and fig. S9). Currents were undetectable for TRPA1.758-OLLAS (fig. S9B). We used TRPA1.755-OLLAS (abbreviated as sTRPA1) in subsequent experiments.

We targeted sTRPA1 to cone photoreceptors of rd1 mice using the same AAV-based approach that was used for rTRPV1 (Fig. 3). To activate the channel, nanorods (λ_{abs} = 915 nm) were conjugated to anti-OLLAS antibodies. Cell membrane expression of sTRPA1 was seen in 50 ± 13% of rd1 cone photoreceptors (table S1, row M, and Fig. 3C). Cones made up 99 ± 0.8% of OLLAS-positive cells (table S1, row N).

To compare rTRPV1 and sTRPA1 sensitivities, we performed behavioral tests in NIR sensor–injected P56-P73 rd1 mice. NIR light of two different intensities cued delayed water appearance for water-restricted, head-fixed animals (Fig. 3D). We evaluated anticipatory lick rates, defined as lick signal after a NIR light flash but before the appearance of water. To measure whether NIR light affects the behavior of wild-type animals, we trained wild-type mice with NIR (915 or 980 nm) and/or visible light (fig. S10A). NIR sensor–injected mice learned to associate NIR light with water within 4 days. At the lower NIR intensity, anticipatory lick rates were similar between control mice and mice with rTRPV1 (table S1, row O) but higher for mice injected with sTRPA1 (table S1, row P, and Fig. 3, E and F). At the higher NIR intensity, rTRPV1 led to higher lick rates compared with control mice (table S1, row Q), but lower than with sTRPA1 (table S1, row R, and Fig. 3, E and F). Behavioral performance of rTRPV1 and sTRPA1 mice was similar to that of wild-type mice trained for 4 days using visible light (table S1, rows S and T, and fig. S10). NIR light neither elicited behavioral responses in wild-type mice nor affected wild-type behavioral responses to visible light (fig. S10C).

To test safety aspects of inducing NIR light sensitivity, we first evaluated the effect of prolonged NIR light exposure on wild-type retinas by immunostaining. NIR light neither activated microglia nor reduced retinal layer thickness, opsin density, or cone density (fig. S11). Second, we tested nanorod biocompatibility with the rd1 retina 80 and 100 days after subretinal injection by immunostaining. Nanorods neither activated microglia, increased apoptosis, nor reduced retinal layer thickness (fig. S12).

Finally, we sought to induce NIR light sensitivity in blind human retinas (Fig. 4). We targeted rTRPV1 to adult human ex vivo retinal explants, in culture for 8 weeks postmortem (Fig. 4B and fig. S13). Retinas lose normal light–evoked activity within 24 hours of isolation (*15*). Using AAV delivery and a CAG promoter, we transduced 2477 ± 889 photoreceptors per square millimeter of human retina (mean ± SD, n = 3 explants) with rTRPV1 (Fig. 4C). Of the rTRPV1-positive cells, 94.5 ± 4.2% were photoreceptors (table S1, row U, and Fig. 4D). To measure whether NIR light drives responses in the human retina, we deposited nanorods (λ_{abs} = 915 nm) over the photoreceptor side. To record calcium signals, we used the fluorescent calcium dye OGB-1 (*16*, *17*). We then performed two-photon calcium imaging of individual neurons in the outer nuclear layer (ONL), inner nuclear layer (INL), and GCL (Fig. 4A). We observed NIR light–induced activation of different human retinal cell classes (Fig. 4, F to H). Most photoreceptors (73%) showed NIR light–evoked increases of calcium signal (Fig. 4H). Some photoreceptors (27%) showed decreases of calcium signal, likely reflecting horizontal cell feedback to NIR light–insensitive photoreceptors (Fig. 4H). In neurons of the INL and GCL, we observed both increases and decreases in calcium signal, indicating activation of excitatory and inhibitory retinal pathways (Fig. 4H). More cells responded in the GCL than in the ONL, reflecting convergent retinal circuit organization (Fig. 4G). Sizes of light-evoked calcium responses were comparable to those in published reports (*16*, *17*).

Here, we described an approach to enable NIR light sensitivity in blind retinas, designed to be compatible with remaining vision (see supplementary text). We used gold nanorods coupled to temperature-sensitive engineered TRP channels to induce NIR light sensitivity in remaining photoreceptor cell bodies of blind mice and in ex vivo human retinas. In mice, NIR light–sensitized photoreceptors activated cortical visual circuits and enabled behavioral responses. By means of distinct nanorods, epitope tags, and TRP channel types, we tuned NIR responses to different wavelengths and to different radiant powers. In the human retina, we reactivated light responses in photoreceptors and their retinal circuits 8 weeks postmortem. Our recordings of NIR light–evoked activity in the postmortem human retina provide not only proof-of-principle for translation but also a model with which the function of human retinal cell types and circuits can be studied.

REFERENCES AND NOTES

1. J. A. Sahel, B. Roska, *Annu. Rev. Neurosci.* **36**, 467–488 (2013).
2. I. Tochitsky, M. A. Kienzler, E. Isacoff, R. H. Kramer, *Chem. Rev.* **118**, 10748–10773 (2018).
3. E. O. Gracheva *et al*., *Nature* **464**, 1006–1011 (2010).
4. E. A. Newman, P. H. Hartline, *Science* **213**, 789–791 (1981).
5. S. A. Stanley *et al*., *Science* **336**, 604–608 (2012).
6. P. P. Joshi, S. J. Yoon, W. G. Hardin, S. Emelianov, K. V. Sokolov, *Bioconjug. Chem.* **24**, 878–888 (2013).
7. Z. Qin, J. C. Bischof, *Chem. Soc. Rev.* **41**, 1191–1217 (2012).
8. M. Lakk *et al*., *Front. Cell. Neurosci.* **12**, 353 (2018).
9. D. A. Ryskamp, S. Redmon, A. O. Jo, D. Križaj, *Cells* **3**, 914–938 (2014).
10. V. Busskamp *et al*., *Science* **329**, 413–417 (2010).
11. J. Jüttner *et al*., *Nat. Neurosci.* **22**, 1345–1356 (2019).
12. S. H. Park *et al*., *J. Immunol. Methods* **331**, 27–38 (2008).
13. C. E. Paulsen, J. P. Armache, Y. Gao, Y. Cheng, D. Julius, *Nature* **520**, 511–517 (2015).
14. L. Zimmermann *et al*., *J. Mol. Biol.* **430**, 2237–2243 (2018).
15. C. S. Cowan *et al*., bioRxiv 703348 [Preprint]. 16 July 2019. https://doi.org/10.1101/703348.
16. K. L. Briggman, T. Euler, *J. Neurophysiol.* **105**, 2601–2609 (2011).
17. T. Baden *et al*., *Nature* **529**, 345–350 (2016).

ACKNOWLEDGMENTS

We thank organ and tissue donors and their families for their generous contributions to science; T. Vögele, J. Sprachta, and P. Blaschke for organizing organ donations; H. Gut and R. Bunker for advice on TRP channel design; J. Jüttner, C. Patino-Alvarez, Ö. Keles, and N. Ledergerber for technical assistance; J. Krol for advice on molecular assays; E. Macé and F. Esposti for assistance with recordings; P. Argast and P. Buchmann for electrical and mechanical engineering in support of the experiments; K. Franke, T. Euler, and Z. Zhao for advice on loading of calcium indicators by electroporation; F. Müller for statistical advice; W. Baehr for sharing of antibodies; D. Dalkara, C. Cepko, and E. Bamberg for plasmids; FMI and NIBR core facilities for their support, especially the microscopy facility, in particular C. Genoud and A. Graff-Meyer for electron microscopy; V. Juvin from SciArtWork for illustrations; and M. Munz and T. Rodrigues for commenting on the manuscript. **Funding:** This work was supported by a Swiss National Science Foundation Synergia grant, a European Research Council advanced grant, a Louis-Jeantet Foundation award, a Swiss National Science Foundation grant, the NCCR "Molecular Systems Engineering" network, a private donation from Lynn and Diana Lady Dougan to B.R., NKFIH 129120 and 2017-1.2.1-NKP-2017-00002 grants to D.H., and a Swiss Academy of Medical Sciences Fellowship to D.N. **Author contributions:** D.N. designed and performed experiments and wrote the paper. C.S.C. performed human retina recordings. R.K.M. performed mouse photoreceptor recordings. Z.R. wrote software. D.G., H.P.N.S., and A.S. contributed to human retina experiments. T.S. performed HEK cell recordings. D.H. performed cortical recordings and analyzed data. B.R. designed experiments and wrote the paper. **Competing interests:** D.N. and B.R. have a patent application on NIR sensors (EP20158285.5). **Data and materials availability:** TRP plasmid materials are available from B.R. under an agreement with the Institute of Molecular and Clinical Ophthalmology Basel. All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6495/1108/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S13
Tables S1 to S4
References (*18*–*40*)

View/request a protocol for this paper from *Bio-protocol*.

21 September 2019; accepted 9 April 2020
10.1126/science.aaz5887

TOPOLOGICAL OPTICS

A fractional corner anomaly reveals higher-order topology

Christopher W. Peterson[1], Tianhe Li[2], Wladimir A. Benalcazar[3], Taylor L. Hughes[2], Gaurav Bahl[4]*

Spectral measurements of boundary-localized topological modes are commonly used to identify topological insulators. For high-order insulators, these modes appear at boundaries of higher codimension, such as the corners of a two-dimensional material. Unfortunately, this spectroscopic approach is only viable if the energies of the topological modes lie within the bulk bandgap, which is not required for many topological crystalline insulators. The key topological feature in these insulators is instead fractional charge density arising from filled bulk bands, but measurements of such charge distributions have not been accessible to date. We experimentally measure boundary-localized fractional charge density in rotationally symmetric two-dimensional metamaterials and find one-fourth and one-third fractionalization. We then introduce a topological indicator that allows for the unambiguous identification of higher-order topology, even without in-gap states, and we demonstrate the associated higher-order bulk-boundary correspondence.

Topological insulators (TIs) are materials with a gapped band structure characterized by quantized quantities, called topological invariants, that are invariant under deformations that preserve both the bulk bandgap and any protective symmetries (*1*, *2*). At a boundary between two materials that have different strong topological invariants—i.e., where a topological invariant changes in space—the bandgap closes, and robust boundary-localized gapless modes appear. Detection of these robust gapless boundary modes is therefore one of the most striking signatures of topological materials.

We focus on two-dimensional (2D) TIs in class AI (spinless and time-reversal symmetric) (*3*). In this class and dimension, no nontrivial strong topological invariants exist (i.e., those protected by particle-hole, chiral, and/or time-reversal symmetry, such as the $\mathbb{Z}_2$ invariant for a quantum spin Hall insulator in class AII), but invariants can be defined if additional spatial symmetries are present. Materials with invariants protected by spatial symmetries are known as topological crystalline insulators (TCIs) (*4*, *5*). We are specifically interested in a recently discovered class of TCIs whose members have gapped boundaries of codimension one but host gapless modes at boundaries with codimension greater than one, i.e., at a boundary of a boundary (*6*–*9*). Because these insulators manifest robust gapless modes at boundaries with higher codimension, they have been termed higher-order topological insulators (HOTIs). We note that spatial symmetries are essential for these HOTIs because they prevent bulk and surface deformations that hybridize and gap out the set of higher-order gapless states.

Only a few naturally occurring HOTIs have been identified (*8*, *9*). Instead, much of the experimental study of HOTIs (primarily d-th order TCIs in d dimensions) has been performed in engineered metamaterials, such as networks of coupled resonators (*10*–*16*), waveguide arrays (*17*, *18*), and photonic or sonic crystals (*19*–*23*). So far, the clearest indicator of higher-order topology in such systems has been the spectroscopic measurement of robust localized corner modes with energies inside the bulk bandgap of 2D (*10*–*14*, *17*–*23*) and 3D (*15*, *16*) HOTIs.

However, there is a fundamental problem with using localized in-gap boundary modes to identify higher-order topology (or, generally, topology protected by spatial symmetries). Spatial symmetries essentially divide a material into symmetric sectors and require that localized modes in each sector are identical. Hence, these symmetries protect the degeneracy of boundary-localized modes but do not restrict their energy (*24*). Additional local symmetries (e.g., chiral symmetry or particle-hole symmetry) can pin the boundary modes to zero energy (midgap) (*24*, *25*), but these symmetries are not actually necessary to protect the higher-order topology, and many lattice models do not support their implementation at all. This implies that the energy of localized boundary modes may reside either in the bulk gap or fully within the bulk bands of a HOTI, depending on the material's details. TIs that fall into the latter case do not host gapless boundary modes within their bulk bandgap and, as such, cannot be distinguished from trivial insulators by their spectrum alone, even with fully open boundary conditions.

This fundamental principle means that HOTIs could be misidentified when their spectra do not exhibit in-gap modes, and it motivates the search for an experimentally measurable indicator of higher-order topology that is protected by only spatial symmetries. It has previously been established that spatial symmetries protect boundary-localized, quantized fractional charge in TCIs (*6*, *24*, *26*–*29*). In this work, we demonstrate that a similar feature in metamaterials—namely, the mode density of the spectral bands—can also be fractionally quantized and can diagnose both first-order and higher-order topology in gapped TCIs. In two dimensions, we term the quantity indicating second-order topology as a fractional corner anomaly (FCA) in the bulk mode density.

We define mode density as the local density of states (DOS) integrated over an entire band, which is equivalent to the charge density of a filled band in an electronic insulator. Unlike charge density, using mode density enables us to study the topology of bands without regard for electronic filling or constraints imposed by charge neutrality. In 2D TCIs, first-order nontrivial topology manifests as an edge-localized fractional mode density σ. For TCIs with only first-order topology, the corner-localized fractional mode density ρ is the sum of the fractional mode densities, σ_1 and σ_2, that respectively manifest at the edges that intersect to form that corner, such that $\rho = \sigma_1 + \sigma_2 \bmod 1$ (modulo operation 1) (*30*). A fractional quantized deviation from this value definitively indicates higher-order topology, such that 2D TCIs without this deviation are not higher-order. To measure this type of higher-order feature in the bulk mode density, we define the FCA ϕ, where

$$\phi = \rho - (\sigma_1 + \sigma_2) \bmod 1 \qquad (1)$$

to capture second-order topology in 2D TCIs. A detailed motivation for the FCA and a generalized proof of this definition are provided in the supplementary materials (*30*).

Notably, although a nonzero FCA does not indicate that corner modes lie within the bulk bandgap, it does indicate the existence of robust topological corner modes somewhere in the spectrum—i.e., either within the bulk bands (or edge bands) or the bandgap. When topological corner modes are not spectrally isolated, the corner modes can generally couple to, and hybridize with, bulk or edge modes, although it was recently shown that in some cases corner modes within a bulk band can act as bound states in the continuum (*31*). However, when spectrally isolated from both the bulk and edge modes, they form the familiar exponentially localized 0D in-gap corner modes (*10*, *12*, *13*, *15*, *16*, *20*). In the

[1]Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [2]Department of Physics and Institute for Condensed Matter Theory, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [3]Department of Physics, The Pennsylvania State University, University Park, PA, USA. [4]Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA.
*Corresponding author. Email: bahl@illinois.edu

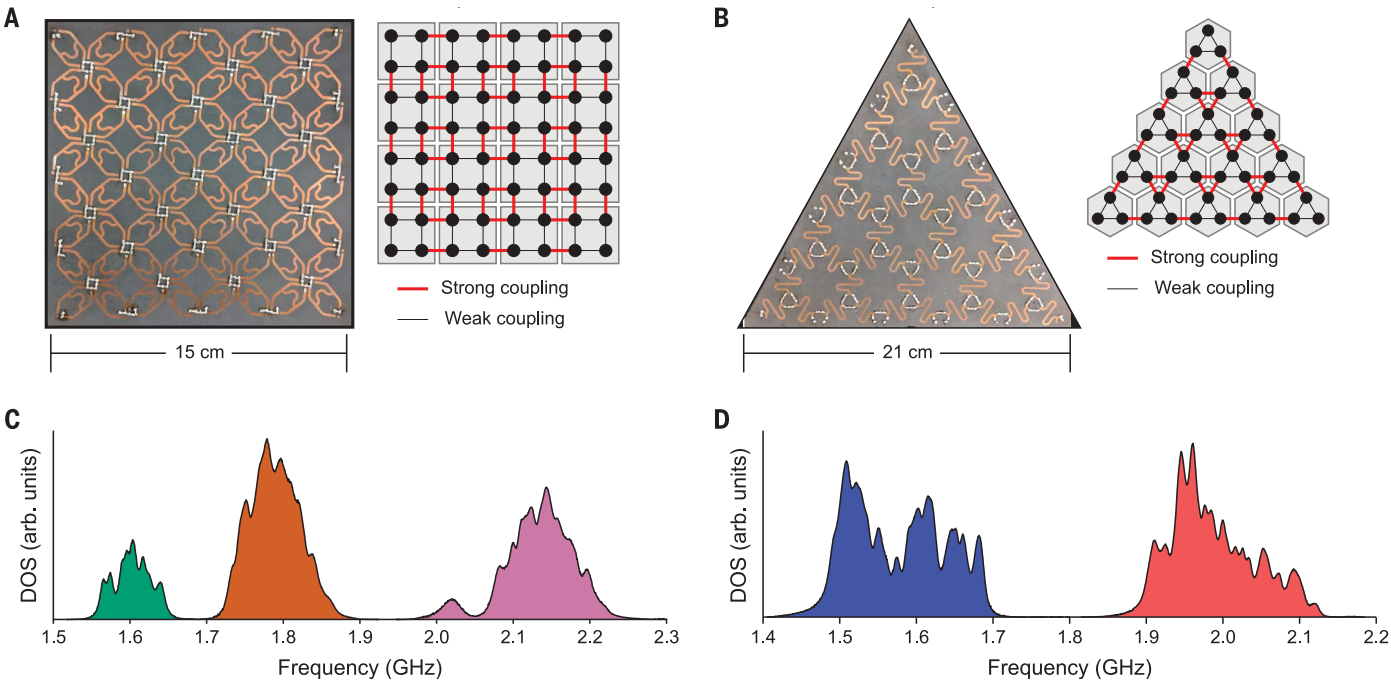


Fig. 1. Fabricated metamaterials and measured spectra. (**A**) Photograph of the experimental resonator array with C_4 symmetry. The schematic on the right illustrates the coupling between resonators. (**B**) Photograph of the experimental resonator array with C_3 symmetry. The schematic on the right illustrates the coupling between resonators. (**C**) Measured DOS spectrum for the resonator array in (A). arb. units, arbitrary units. (**D**) Measured DOS spectrum for the resonator array in (B).

next section, we experimentally measure a nonzero FCA in insulators where the corner modes are hybridized with bulk modes. Simulation results, detailed in the supplementary materials (*30*), show that the energy of topological corner modes can be tuned into, and even fully across, the bulk bandgaps (and any edge bandgaps) when a localized potential is applied to only the corner unit cells. We then demonstrate that these corner modes can be spectrally isolated and exponentially localized by deformation of the corner unit cell.

To observe the FCA experimentally, we constructed two rotationally symmetric TI metamaterials in microwave-frequency coupled resonator arrays. We chose to test two insulators with different symmetries because the quantization of the fractional mode density and FCA depends on the rotation symmetry group (*24*). The first insulator, shown in Fig. 1A, is on a square lattice with C_4 symmetry, and the second insulator, shown in Fig. 1B, is on a kagome lattice with C_3 symmetry. We first found the spectral DOS of both metamaterials by means of reflection measurements; see the supplementary materials (*30*) for details of the measurement technique. The measured spectrum of the C_4-symmetric insulator, shown in Fig. 1C, has three distinct bands. The measured spectrum of the C_3-symmetric insulator, shown in Fig. 1D, has two bands. Neither of these insulators have in-gap modes, so from the spectra alone it is not possible to tell whether either metamaterial is topologically nontrivial. However, as we will show, both are in fact nontrivial, and the intrinsic chiral symmetry breaking causes the edge and corner modes to lie within the bulk bands (*30*).

Next, we calculate the mode density of the measured bands by integrating the local DOS in each unit cell over their respective frequency ranges, as shown for both insulators in Fig. 2. The mode density of the C_4-symmetric insulator is shown in Fig. 2A and has several important features on which we will focus. First, we find that the resonators in the bulk unit cells are excited in all three bands, which indicates that this insulator nominally has three bulk bands. We observe that the total mode density of these bands in each sector is approximately equal, which demonstrates that this insulator has approximate C_4 rotation symmetry with a small amount of symmetry-breaking disorder from fabrication imperfections. As expected, the mode density in the bulk unit cells, designated by $\mu^{(n)}$ (where the superscript indicates C_n symmetry), is always an integer. Specifically, $\mu_{1,3}^{(4)} \approx 1$ (where the subscript indicates the band number) and $\mu_2^{(4)} \approx 2$, which indicates that band 2 is a twofold degenerate band whereas bands 1 and 3 are nondegenerate.

Most importantly, we find a nonzero fractional mode density in the edge and corner unit cells. Because the bandgaps are relatively large compared with the width of the bands (i.e., the system has a very short correlation length), the entirety of the fractional mode density is tightly localized within the boundary unit cells. For bands 1 and 3, the fractional mode density in the edge unit cells is $\sigma_{1,3}^{(4)} \approx \frac{1}{2}$, and in the corner unit cells it is $\rho_{1,3}^{(4)} \approx \frac{1}{4}$. In the twofold degenerate band 2, these fractions are doubled. This doubling occurs because band 2 is equivalent to two copies of band 1, in the sense that the Wannier representation of the orbitals in band 2 has two Wannier centers at Wyckoff position *b*, whereas band 1 has only one at position *b* (*30*); each Wannier center contributes equal fractional mode density to the boundary. The approximate fractions written above are obtained by rounding to the nearest quarter, as we anticipate that C_4 symmetry quantizes mode density in fractions of one-fourth (*24*).

We can now extract the FCA ϕ for each bulk band using the mode density data in Fig. 2A. Because the system is near a zero–correlation-length limit, we find ϕ using the simple formula $\phi = \rho - 2\sigma$, where ρ is the fractional mode density of the corner unit cell and σ is the fractional mode density of the edge unit cells (because of C_4 symmetry, all edges are expected to be identical). Here, because there is a small amount of unavoidable disorder in the experiment (which slightly breaks C_4 symmetry), we average over all the edges to find σ and over all the corners to find ρ,

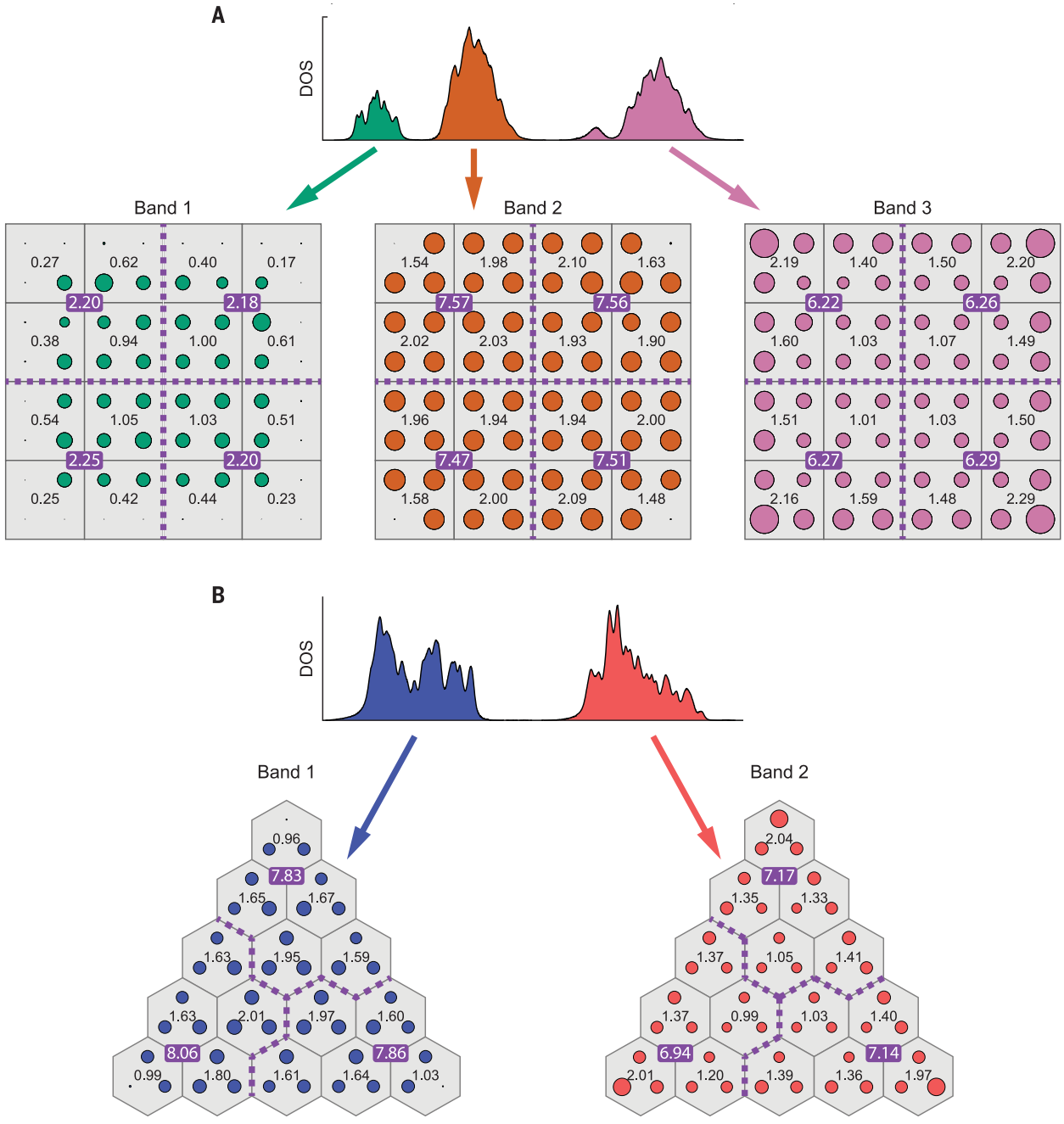


Fig. 2. Spatial distribution of bands. (**A**) Measured mode density for the C_4-symmetric insulator. The mode density for each band is shown separately. Each filled circle represents a resonator, with the area of the circle corresponding to the measured mode density of that resonator in that band. The total mode density of each unit cell is shown with black text, and the total mode density of each sector is shown in purple with white text. (**B**) Same as (A) but for the C_3-symmetric insulator.

such that $\phi_1^{(4)}$—i.e., the FCA for band 1 in the C_4-symmetric metamaterial—is

$$\phi_1^{(4)} = \rho_1^{(4)} - 2\sigma_1^{(4)} = 0.23 - 0.98 = 0.25 \approx \frac{1}{4} \quad (2)$$

A similar calculation can be carried out for the other bulk bands, giving $\phi_2^{(4)} = 0.55 \approx \frac{1}{2}$ for band 2 and $\phi_3^{(4)} = 0.19 \approx \frac{1}{4}$ for band 3. The sum of the FCA over all the bulk bands is always an integer (here rounding gives $\phi_1^{(4)} + \phi_2^{(4)} + \phi_3^{(4)} = 0.99$), and $\phi_2^{(4)}$ is twice $\phi_1^{(4)}$, as expected.

Next, we move on to the mode density for the C_3-symmetric system, shown in Fig. 2B. For this material, we again find that the bulk unit cells have integer mode density, $\mu_1^{(3)} \approx 2$ (a twofold degenerate band) and $\mu_1^{(3)} \approx 1$. As in the previous system, the total mode density of these bands in each sector is approximately equal. The edge unit cells have a fractional mode density of $\sigma_1^{(3)} \approx \frac{2}{3}$ in band 1 and $\sigma_2^{(3)} \approx \frac{1}{3}$ in band 2. Here, the approximate fractions are obtained by rounding to the nearest third because this system is C_3 symmetric (*24*). Notably, although this material does not have a fractional mode density in the corner unit cells, the FCA indicator is nonzero, $\phi_1^{(3)} = 0.70 \approx \frac{2}{3}$, and $\phi_2^{(3)} = 0.30 \approx \frac{1}{3}$. For comparison, in the supplementary materials (*30*) we also present experimental measurements of the FCA for a trivial insulator and show that $\phi^{(3),\,\mathrm{triv}} \approx 0$ for all bands.

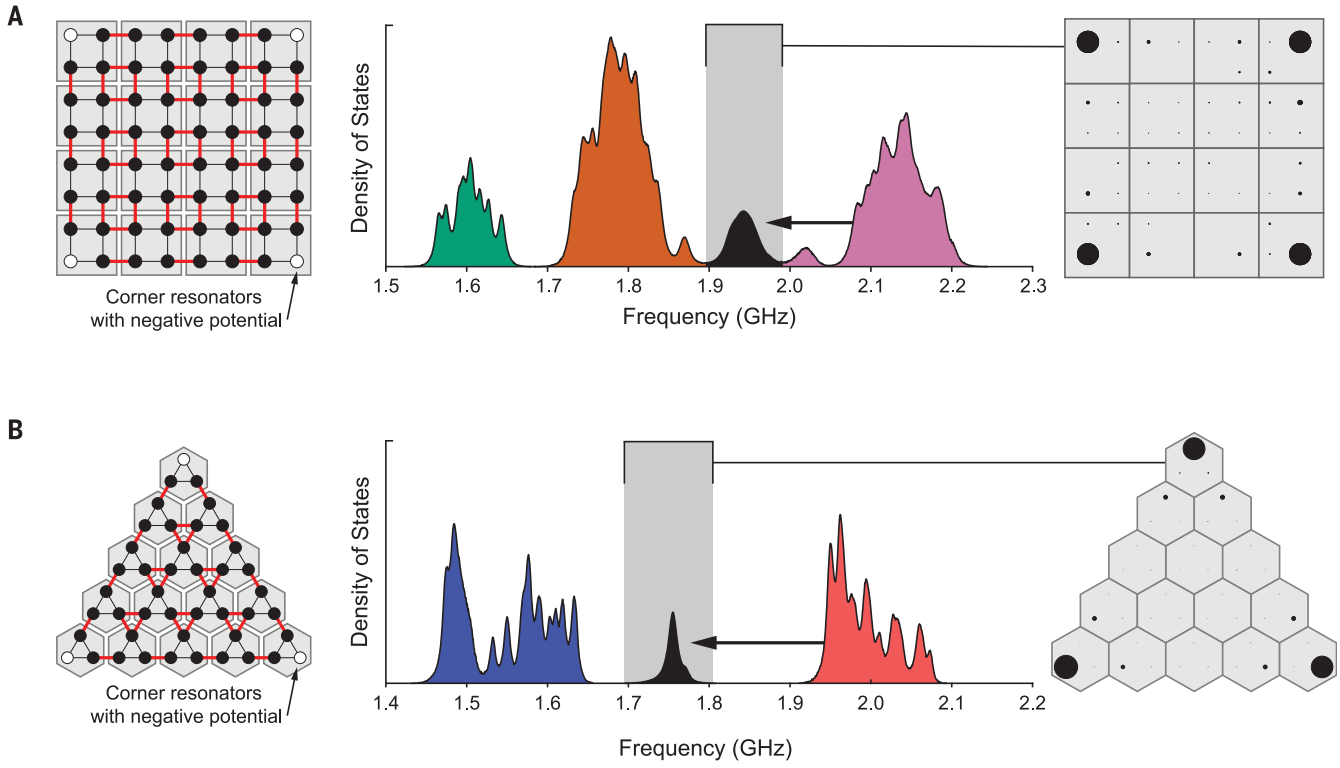


Fig. 3. Pulling topological modes into the gap. (**A**) The schematic on the left shows where a small, negative on-site potential is applied (white circles). The measured DOS has modes within the bulk bandgap, which are spatially localized to the corners and confined to one sublattice. (**B**) Same as (A) but for the C_3-symmetric insulator.

The nonzero FCA in both metamaterials indicates that they are indeed HOTIs, and we have argued above that they should host second-order topological modes at their corners. Because we have not observed these expected topological modes within the bulk bandgap, we can estimate their spectral location by finding the band in which the corner resonators are most strongly excited. In the C_4-symmetric system, the corner resonators, around which the second-order topological modes are expected to exist, are mainly excited in band 3, which indicates that the corner modes lie in this band. Moreover, this implies that we can spectrally localize these modes by slightly lowering the resonance frequency of the corner resonators. As illustrated in Fig. 3A, we applied a small negative potential to the corner resonators using a capacitor connected to ground, which decreases the electrical length (and thus the resonance frequency) of the corner resonators. When the potential is applied, we observe topological modes moving into the bandgap between bands 2 and 3 and becoming exponentially localized to the corner and confined to one sublattice.

In the C_3-symmetric system, the corner resonators are only excited in band 2, which again indicates that the energy of the corner modes is too high and should be lowered to bring the modes into the bandgap. We pull these modes into the bandgap by similarly applying a small negative potential to the corners, as illustrated in Fig. 3B. The topological modes are observed to spectrally localize within the bandgap and spatially localize to the corners with confinement on one sublattice. We also conducted a similar experiment on a trivial insulator (*30*) and found that localized corner modes cannot be isolated within the bulk bandgap, regardless of the applied potential strength.

The definition of the FCA can also be extended beyond two dimensions to identify d-th–order topology in fully gapped d-dimensional insulators. For example, the FCA for third-order TCIs in three dimensions is

$$\phi = \sum_{i,j} (\delta - \rho_j - \sigma_i) \bmod 1 \qquad (3)$$

where δ is corner-localized fractional mode density, ρ_j is hinge-localized fractional mode density, and σ_i is surface-localized fractional mode density. Because this indicator captures fundamental topological features that are protected by spatial symmetries, we expect that it can assist with the experimental identification of materials with higher-order topology, which could otherwise be misidentified by only searching for in-gap corner modes. From a practical perspective, focusing on bulk-derived fractional mode density instead of localized modes could simplify experimental confirmations of novel TIs, which often use ad hoc supplementary boundary elements (e.g., auxiliary resonators or loading capacitors) to spectrally shift topological modes into the bandgap (*22*–*25*, *28*, *29*, *31*).

REFERENCES AND NOTES

1. X.-L. Qi, S.-C. Zhang, *Rev. Mod. Phys.* **83**, 1057–1110 (2011).
2. M. Z. Hasan, C. L. Kane, *Rev. Mod. Phys.* **82**, 3045–3067 (2010).
3. A. P. Schnyder, S. Ryu, A. Furusaki, A. W. Ludwig, *Phys. Rev. B* **78**, 195125 (2008).
4. R.-J. Slager, A. Mesaros, V. Juričić, J. Zaanen, *Nat. Phys.* **9**, 98–102 (2013).
5. Y. Ando, L. Fu, *Annu. Rev. Condens. Matter Phys.* **6**, 361–381 (2015).
6. W. A. Benalcazar, B. A. Bernevig, T. L. Hughes, *Science* **357**, 61–66 (2017).
7. Z. Song, Z. Fang, C. Fang, *Phys. Rev. Lett.* **119**, 246402 (2017).
8. F. Schindler *et al.*, *Sci. Adv.* **4**, eaat0346 (2018).
9. F. Schindler *et al.*, *Nat. Phys.* **14**, 918–924 (2018).
10. C. W. Peterson, W. A. Benalcazar, T. L. Hughes, G. Bahl, *Nature* **555**, 346–350 (2018).
11. S. Imhof *et al.*, *Nat. Phys.* **14**, 925–929 (2018).
12. H. Xue, Y. Yang, F. Gao, Y. Chong, B. Zhang, *Nat. Mater.* **18**, 108–112 (2019).
13. X. Ni, M. Weiner, A. Alù, A. B. Khanikaev, *Nat. Mater.* **18**, 113–120 (2019).
14. S. Mittal *et al.*, *Nat. Photonics* **13**, 692–696 (2019).
15. H. Xue *et al.*, *Phys. Rev. Lett.* **122**, 244301 (2019).
16. M. Weiner, X. Ni, M. Li, A. Alù, A. B. Khanikaev, *Sci. Adv.* **6**, eaay4166 (2020).
17. J. Noh *et al.*, *Nat. Photonics* **12**, 408–415 (2018).
18. A. El Hassan *et al.*, *Nat. Photonics* **13**, 697–700 (2019).
19. M. Serra-Garcia *et al.*, *Nature* **555**, 342–345 (2018).
20. X. Zhang *et al.*, *Nat. Phys.* **15**, 582–588 (2019).
21. Z.-K. Lin, H.-X. Wang, Z. Xiong, M.-H. Lu, J.-H. Jiang, Nonsymmorphic Quadrupole Topological Insulators in Sonic Crystals. arXiv:1903.05997 [cond-mat.str-el] (14 March 2019).
22. B. Y. Xie *et al.*, *Phys. Rev. Lett.* **122**, 233903 (2019).
23. X. D. Chen *et al.*, *Phys. Rev. Lett.* **122**, 233902 (2019).
24. W. A. Benalcazar, T. Li, T. L. Hughes, *Phys. Rev. B* **99**, 245151 (2019).
25. J. Ahn, S. Park, B.-J. Yang, *Phys. Rev. X* **9**, 021013 (2019).

26. W. P. Su, J. R. Schrieffer, A. J. Heeger, *Phys. Rev. Lett.* **42**, 1698–1701 (1979).
27. D. Vanderbilt, R. D. King-Smith, *Phys. Rev. B* **48**, 4442–4455 (1993).
28. R. Resta, *Rev. Mod. Phys.* **66**, 899–915 (1994).
29. G. van Miert, C. Ortix, *Phys. Rev. B* **98**, 081110 (2018).
30. Additional analyses are available as supplementary materials.
31. W. A. Benalcazar, A. Cerjan, Bound states in the continuum of higher-order topological insulators. arXiv:1908.05687 [cond-mat.str-el] (15 August 2019).

ACKNOWLEDGMENTS

The authors would like to thank J. T. Bernhard for access to the resources at the University of Illinois at Urbana-Champaign Electromagnetics Laboratory, and they thank Q. Wang and W. Jiang for assisting with the fabrication and measurement of the microwave circuits. **Funding:** This project was supported by the U.S. National Science Foundation (NSF) Emerging Frontiers in Research and Innovation (EFRI) grant EFMA-1627184. C.W.P. acknowledges support from the NSF Graduate Research Fellowship. G.B. acknowledges support from the U.S. Office of Naval Research (ONR) Director for Research Early Career Grant N00014-17-1-2209. T.L., W.A.B., and T.L.H. also thank the U.S. NSF under grant DMR-1351895. **Author contributions:** C.W.P. designed and fabricated the microwave circuits, performed the microwave simulations and experimental measurements, and produced the experimental figures. T.L. and W.A.B. guided the TI design and performed the theoretical calculations. T.L.H. and G.B. supervised all aspects of the project. All authors jointly wrote the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6495/1114/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S4
References (*32*–*45*)

2 January 2020; accepted 14 April 2020
10.1126/science.aba7604

PALEOECOLOGY

Thresholds of mangrove survival under rapid sea level rise

N. Saintilan[1]*, N. S. Khan[2,3], E. Ashe[4], J. J. Kelleway[5,6], K. Rogers[5,6], C. D. Woodroffe[5,6], B. P. Horton[7,8]

The response of mangroves to high rates of relative sea level rise (RSLR) is poorly understood. We explore the limits of mangrove vertical accretion to sustained periods of RSLR in the final stages of deglaciation. The timing of initiation and rate of mangrove vertical accretion were compared with independently modeled rates of RSLR for 78 locations. Mangrove forests expanded between 9800 and 7500 years ago, vertically accreting thick sequences of organic sediments at a rate principally driven by the rate of RSLR, representing an important carbon sink. We found it very likely (>90% probability) that mangroves were unable to initiate sustained accretion when RSLR rates exceeded 6.1 millimeters per year. This threshold is likely to be surpassed on tropical coastlines within 30 years under high-emissions scenarios.

The rate of relative sea level rise (RSLR) in tropical and subtropical locations is projected to accelerate from current trends of ~3.4 mm year^{-1} to a mean estimate of ~5 mm year^{-1} under low-emissions scenarios and ~10 mm year^{-1} under high-emissions scenarios by 2100 (*1*, *2*). Modeling of feedbacks between RSLR, vertical accretion, root mass formation, and tidal marsh and mangrove vulnerability under sustained high rates of RSLR is vital for the survival of these ecologically and economically important coastal environments (*2*). Some tidal marshes have been projected to survive RSLR of more than 10 mm year^{-1} where supported by high available suspended sediment concentrations (*3*), on the basis of accretion data that span annual to decadal time scales. Reconstructions using paleoenvironmental proxies, however, have suggested that UK marshes were vulnerable to retreat at RSLR of 7 mm year^{-1} (*4*).

Mangroves grow in sheltered intertidal environments that are exposed to the effects of RSLR (*5*). They support among the highest rates of carbon burial of all ecosystems (*6*), and a growing body of evidence suggests that this efficiency is enhanced by RSLR (*7*). However, empirical data on mangrove response to high rates of RSLR are lacking, given the limited observation period (1 to 16 years) of real-time measurements (*5*).

Here, we assess the mangrove response to RSLR from the paleorecord of mangrove vertical accretion preserved in the sedimentary archives of continental shelves and coastal lowlands. Mangrove forests have established and drowned in association with variability in the rate of RSLR after the onset of deglaciation ~26,000 to 20,000 years ago (*8*). Before the Holocene, long-term RSLR rates of >12 mm year^{-1} (*8*) exceeded the capacity of mangroves to maintain position in situ through vertical accretion, and mangroves were displaced landward (*9*), with few exceptions (*10*, *11*). During the early to mid-Holocene, RSLR slowed in association with the final phase of deglaciation of the Laurentide ice sheet (*12*). The rate of RSLR varied across the globe owing to glacio-isostatic adjustment, the response of the solid Earth and gravity field to ice mass redistribution during a glacial cycle (*13*). Far-field sites, those distal from regions of former ice sheet extent and incorporating most of the tropics, exhibit a decline in the rate of RSLR (*14*) from 10,000 to 7000 years ago (phase 1 in Fig. 1A) to a mid-Holocene highstand, followed by stable or falling sea level to near the current position from ~6000 years ago to the present (phase 2 in Fig. 1A). By contrast, at intermediate-field sites closer to centers of glaciation (for example, the Gulf of Mexico and the Caribbean), sea level rose continuously throughout the Holocene owing to isostatic response to melting of the Laurentide ice sheet. These coastlines have experienced a decelerating rate of RSLR from 10,000 years ago to the present (Fig. 1B and figs. S5 and S6).

The spatially variable deceleration of RSLR from 10,000 to 8000 years ago across tropical and subtropical latitudes coincided with the initiation of vertically continuous, organic-rich mangrove sediments several meters thick, as rising seas flooded shallow continental shelves (phase 1 in Fig. 1, A and B) (*15*). This deceleration provides an opportunity to explore whether (i) the rate of mangrove vertical accretion responds to changes in RSLR; (ii) ice sheet proximity (intermediate versus far field), geomorphic setting, or tidal range constrains the capacity of mangroves to accrete in relation to RSLR; (iii) upper thresholds of mangrove vertical accretion can be detected; and (iv) mangrove development and vertical accretion correspond in timing to changes in the global atmospheric carbon budget.

We present empirical data from 122 reconstructions of the timing and rate of mangrove vertical accretion associated with Holocene RSLR in cores collected from 78 tropical and subtropical locations [Fig. 2; (*14*)]. We independently estimate rates of RSLR for each of the 78 locations before and for the duration of mangrove vertical accretion from a glacio-isostatic adjustment model using an ensemble of different Earth parameters (*16*) (Fig. 3 and figs. S5 and S6).

Slowing RSLR during the early to mid-Holocene coincided with the initiation of extensive mangrove forests (Figs. 1A, phase 1, and 3, A and B). Our analysis suggests that sustained mangrove vertical accretion began across far-field regions (Africa, Asia, Australasia, and South America) at ~10,000 to 8000 years ago and intermediate-field regions (Caribbean and Gulf of Mexico) at ~8000 to 6000 years ago (Fig. 3 and figs. S1 and S2). Data were discriminated on the basis of ice sheet proximity and geomorphic setting and differentiated by tidal regime to explore differences in the timing of initiation and rates of vertical accretion (*14*). The intermediate- and far-field classifications, defined as a function of a location's proximity to areas of major ice sheet retreat during the last deglaciation, act as a surrogate variable for the temporal pattern of RSLR rates through the Holocene, which in turn influences the availability of accommodation space within which mangroves can accrete vertically. In a generalized linear model of several variables, only the proximity to former ice sheets proved to have a significant relationship with accretion rates (*14*).

We found a strong relationship ($p < 0.001$, generalized linear model) between rates of mangrove vertical accretion and RSLR rate across all sites (51% of the variation in the accretion rate can be explained by the RSLR variation; Fig. 3). Mangrove vertical accretion first initiated ~9800 years ago in the Ganges-Brahmaputra River Delta, as RSLR decreased from ~9 to ~6 mm year^{-1}, and continued for ~650 years at high rates until replaced by subtidal deposits as RSLR increased again to >7 mm year^{-1} (*14*, *17*). Elsewhere in far-field locations, mangrove vertical accretion initiated as RSLR decreased below ~7 mm year^{-1} starting 8800 years ago (Fig. 3). Rates of vertical accretion of >6 mm year^{-1} were sustained for more than 1000 years in mangrove forests in Australia, Africa, South America, Central America, and Asia (Fig. 2 and data S1), irrespective of geomorphic setting (fig. S1).

Mangrove vertical accretion initiated in intermediate-field regions later than in far-field locations, first in Belize ~7800 years ago as

[1]Department of Earth and Environmental Sciences, Macquarie University, Macquarie Park, NSW, Australia. [2]Department of Earth Sciences, University of Hong Kong, Hong Kong, China. [3]Swire Institute of Marine Science, University of Hong Kong, Hong Kong, China. [4]Department of Earth and Planetary Sciences, Rutgers University, Piscataway NJ, USA. [5]School of Earth, Atmospheric, and Life Sciences, University of Wollongong, Wollongong, NSW, Australia. [6]Geoquest Research Centre, University of Wollongong, Wollongong, NSW, Australia. [7]Asian School of Environment, Nanyang Technological University, Singapore. [8]Earth Observatory of Singapore, Nanyang Technological University, Singapore.
*Corresponding author. Email: neil.saintilan@mq.edu.au

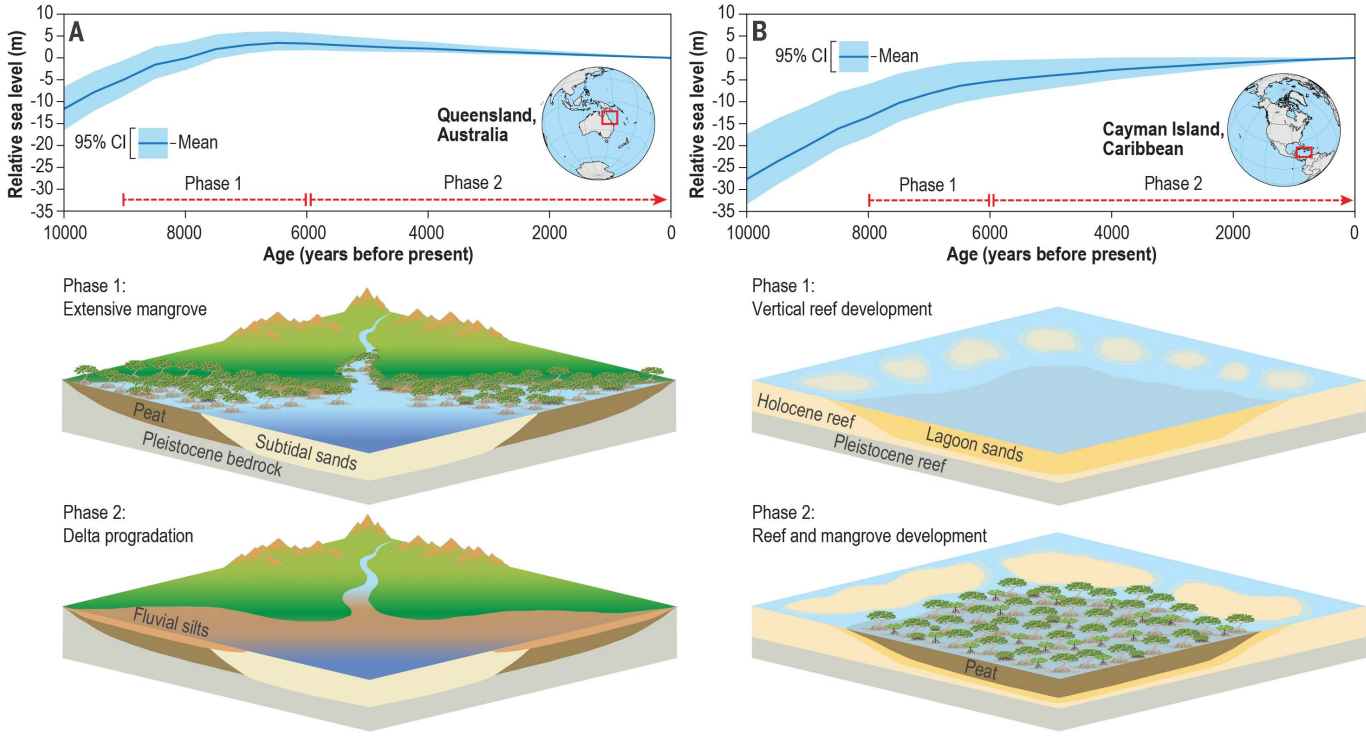
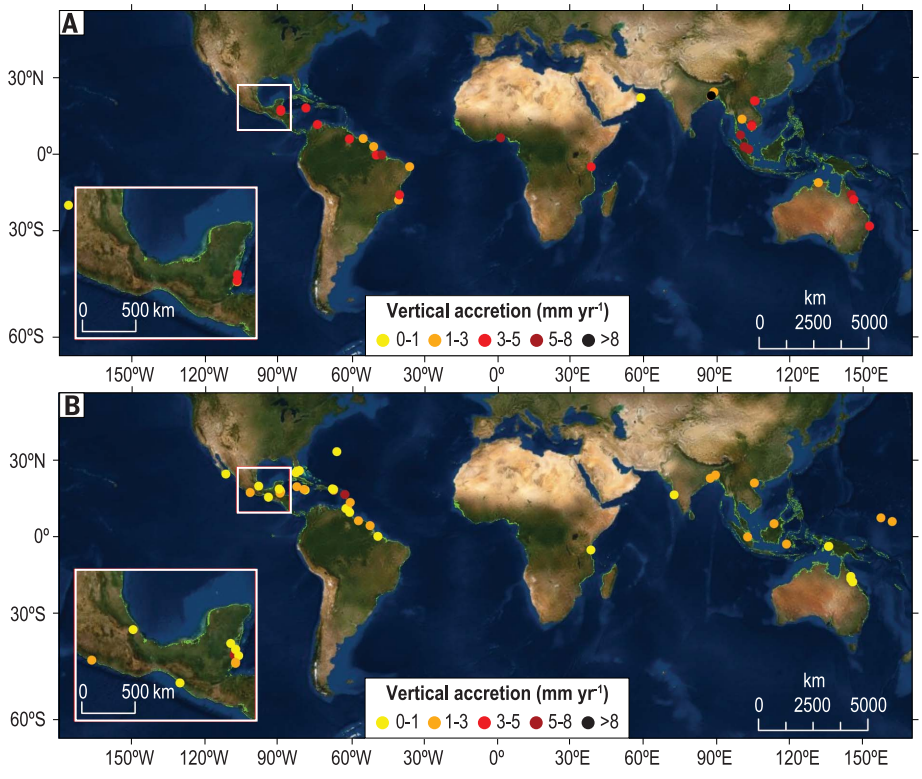


Fig. 1. Holocene RSLR and associated mangrove development in far-field estuarine and intermediate-field calcareous settings. (**A**) Far-field estuarine settings. (**B**) Intermediate-field calcareous settings. As the rate of RSLR decelerated, development of organic-rich mangrove sediments was initiated, which facilitated maintenance of vertical position with respect to sea level, forming extensive mangrove forests. When RSLR stabilizes, ongoing sedimentation infills estuaries, building deltas and replacing mangrove with freshwater wetlands, terrestrial forest, or saltpan, depending on climatic setting. CI, credible interval.

Fig. 2. Sampling locations and corresponding rates of mangrove vertical accretion. (**A** and **B**) Mangrove sedimentary units initiating between 10,000 and 6500 years ago (A) and mangrove sedimentary units initiating after 6500 years ago (B). Contemporary mangrove shorelines are indicated in bright green, representing mangrove distribution in 2012 (*30*).



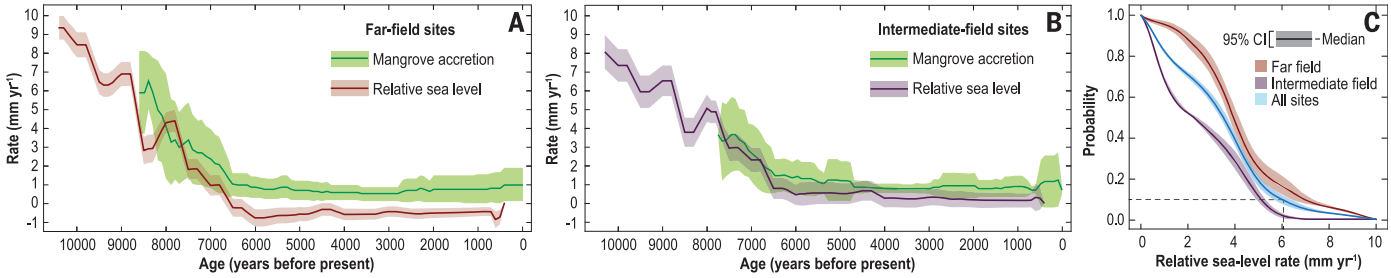


Fig. 3. Rates of mangrove vertical accretion and RSLR. (**A** and **B**) Accretion rates are derived from the depth of mangrove organic sediment between calibrated ^{14}C dates in individual cores, and rates of RSLR (median and 95% CI) were derived from the glacio-isostatic adjustment (GIA) model ensemble for the same sites for far-field (A) and intermediate-field (B) locations. (**C**) The probability of initiation of sustained mangrove accretion at or above associated rates of RSLR across all sites. The dotted line shows the RSLR rate at which there is a 10% probability that mangroves can initiate accretion and/or, conversely, that there is a 90% probability that they are unable to initiate accretion.

Table 1. Probability that mangroves are unable to initiate sustained vertical accretion at rates of RSLR. See (*14*). RSLR rates at all sites (global) and intermediate- and far-field sites at which it is very likely (>90% probability) and extremely likely (>95% probability) that mangroves are unable to initiate sustained accretion and the associated 95% uncertainty interval (UI). For example, at all sites, there is a 94.3 to 95.4% probability (95% UI) that mangroves are unable to initiate at rates that exceed 7.6 mm year^{-1}.

Sites	RSLR rates above which accretion did not initiate (very likely: 90% median probability)		RSLR rates above which accretion did not initiate (extremely likely: 95% median probability)	
	RSLR rate (mm year^{-1})	Uncertainty interval (%)	RSLR rate (mm year^{-1})	Uncertainty interval (%)
Global	6.1	(88.1–92.8)	7.6	(94.3–95.4)
Intermediate-field (Gulf of Mexico, Caribbean)	5.2	(88.1–92.8)	5.7	(93.7–96.9)
Far-field	7.1	(87.3–91.6)	8.8	(94.4–95.8)

RSLR fell below ~5 mm year^{-1} and in other Caribbean and Gulf of Mexico sites between 7500 and 6000 years ago (Figs. 2 and 3 and table S1). This later initiation, in comparison to far-field regions, may be related to the limited allochthonous sediment supply available in the carbonate (reef) settings, although in Belize, rates of accretion of more than 6 mm year^{-1} were observed at two sites (data S1), driven by strong authochthonous inputs.

We used a Bayesian framework to estimate the probability of initiation of mangrove accretion conditional on rates of RSLR within the Holocene dataset (*14*). The empirical Holocene data (data S1), which span a wide range of geomorphic settings and tidal regimes (fig. S2), suggest that when RSLR rates exceed 6.1 and 7.6 mm year^{-1}, respectively, mangroves are very likely (>90% probability) and extremely likely (>95% probability) to be unable to initiate sustained accretion. We found lower RSLR thresholds for intermediate-field sites and higher thresholds for far-field sites (Table 1).

Our database also reveals spatial variability in the duration of mangrove accretion (figs. S1 and S2). In only 3 of 50 far-field locations was there indication of drowning of mangroves during a marine transgression (i.e., mangrove sediments overlain by tidal or subtidal deposits; data S1). In most far-field locations (30 of 50), accretion and progradation of the fluvial delta led to the replacement of mangrove by saltmarsh (characteristic of upper intertidal elevations), freshwater wetland, or terrestrial forest, often by the mid-Holocene when relative sea level stabilized (phase 2 in Fig. 1A and data S1). For this reason, contemporary mangrove extent is highly contracted compared with the early- and mid-Holocene mangrove development in many major river deltas such as the Ord River, Australia; the Red River, Vietnam; and the Mekong River, Vietnam and Cambodia (*15*, *18*–*20*). Mangrove accretion persisted significantly longer at intermediate-field locations (fig. S2) because accommodation space was enhanced by glacio-isostatic adjustment.

The extensive development of mangrove environments under RSLR has exerted an influence on global carbon cycles over geologic time scales (*21*), including, we suggest, during the early to mid-Holocene (Fig. 4). Carbon balance modeling based on stable carbon isotope signatures suggests that the 5 parts per million by volume reduction in atmospheric CO_2 in the early Holocene was driven by increases in the uptake of carbon by the land biosphere on the order of 290 Pg C (*22*, *23*) before 7000 years ago. The timing and volume of this uptake have been attributed to the northward expansion of boreal vegetation after ice sheet retreat [~110 Pg C (*24*)] and organic soil development [180 Pg C (*23*)], of which circum-Arctic peatlands may have sequestered 20 to 60 Pg C based on the depth of peat formation at the time (*24*). We conservatively estimate a somewhat larger contribution (~85 Pg C) from mangrove carbon sequestration and burial over the period 8600 to 6000 years ago (*14*). The extensive development of mangrove forests over this period largely replaced methanogenic environments {freshwater wetlands and floodplains [data S1; (*25*)]} and corresponds to declining rates of methane emission, particularly in the Southern Hemisphere (*26*). The $\delta^{13}CH_4$ signals in the Southern Hemisphere for this period show a 1.5 per mil depletion consistent with a replacement of vegetation using the C_4 photosynthetic pathway (tropical grasslands and saltmarsh adapted to low atmospheric CO_2), with mangroves using the C_3 pathway (*26*, *27*).

As RSLR increases in the 21st century, an increase in the rate of mangrove vertical accretion, coupled with landward expansion as sea level rises, can be expected to drive increases in the rate of carbon sequestration and preservation in mangrove environments, providing a negative feedback on radiative forcing, as suggested more broadly for coastal wetlands (*7*). Although

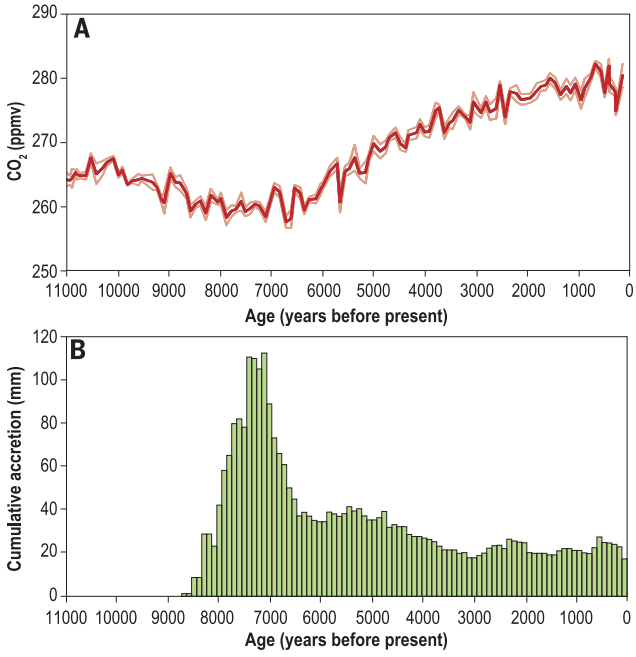


Fig. 4. Timing of mangrove vertical accretion in relation to greenhouse gas concentrations. (**A** and **B**) Atmospheric CO_2 concentrations [from Antarctic ice cores (*23*)] (A) and mangrove organic soil formation (B) [the sum of all observations spanning each century weighted by the vertical accretion rate of each observation (*14*)]. Light colored curves in (A) represent ± 1 standard error for CO_2 (B), as presented in the original datasets. ppmv, parts per million by volume.

our results demonstrate that accretion in mangroves increases in response to RSLR, we found that it was very likely (>90% probability) that mangroves were unable to initiate sustained accretion when RSLR rates exceeded 6.1 mm $year^{-1}$ in any but the most sediment-laden settings. RSLR is projected to remain below 5 mm $year^{-1}$ under low-emissions scenarios [Representative Concentration Pathway (RCP) 2.6] throughout the 21st century (*1*). However, RSLR is expected to exceed 5 mm $year^{-1}$ by 2030 and 7 mm $year^{-1}$ by 2050 under high-emissions scenario RCP8.5 in low-latitude mangrove settings where rates of RSLR are expected to be higher than the global average (*1*, *2*).

Where a deficit commences between vertical accretion and RSLR, time to submergence will be a function of the position of the mangrove within the tidal frame. In settings of low tidal range, mangroves are more likely to be situated at elevations close to the threshold of submergence from the outset. In settings of high tidal range, mangroves are more likely to be situated at elevations well above this threshold and tolerate a deficit between the rates of accretion and RSLR for decades to centuries (*5*). Geomorphic setting will also influence vulnerability to submergence, because allochthonous sediment contributions in tide- and river-dominated estuaries may provide an elevation subsidy not available in environments receiving low sediment supply, such as coral reefs. In this context, sediment retention in catchments affected by water resource development (i.e., trapped behind dams) and local sediment controls may decrease mangrove resilience to RSLR in river estuaries (*5*). The natural response of mangrove encroachment across flooded coastal lowlands is therefore the main determinant of future extent (*28*), although this is already greatly impeded by coastal developments along many coastlines (*29*). Our findings therefore emphasize the importance of (i) mitigating the magnitude of rapid RSLR and (ii) ensuring that coastal adaptation measures allow for the expansion of mangrove across coastal lowlands.

REFERENCES AND NOTES

1. J. A. Church *et al.*, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, T. F. Stocker *et al.*, Eds. (Cambridge Univ. Press, 2013), chap. 13.
2. R. E. Kopp *et al.*, *Earths Futur.* **2**, 383–406 (2014).
3. M. L. Kirwan, S. Temmerman, E. E. Skeehan, G. R. Guntenspergen, S. Fagherazzi, *Nat. Clim. Chang.* **6**, 253–260 (2016).
4. B. P. Horton *et al.*, *Nat. Commun.* **9**, 2687 (2018).
5. C. E. Lovelock *et al.*, *Nature* **526**, 559–563 (2015).
6. D. C. Donato *et al.*, *Nat. Geosci.* **4**, 293–297 (2011).
7. K. Rogers *et al.*, *Nature* **567**, 91–95 (2019).
8. K. Lambeck, H. Rouby, A. Purcell, Y. Sun, M. Sambridge, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 15296–15303 (2014).
9. C. D. Woodroffe, *Mar. Freshw. Res.* **69**, 917–932 (2018).
10. T. J. Hanebuth, H. K. Voris, Y. Yokoyama, Y. Saito, J. I. Okuno, *Earth Sci. Rev.* **104**, 92–110 (2011).
11. W. A. Nicholas *et al.*, *Cont. Shelf Res.* **83**, 129–142 (2014).
12. N. S. Khan *et al.*, *Curr. Clim. Change Rep.* **1**, 247–262 (2015).
13. W. E. Farrell, J. A. Clark, *Geophys. J. Int.* **46**, 647–667 (1976).
14. Materials and methods are available as supplementary materials.
15. C. D. Woodroffe, B. G. Thom, J. Chappell, *Nature* **317**, 711–713 (1985).
16. S. Dendy, J. Austermann, J. R. Creveling, J. X. Mitrovica, *Quat. Sci. Rev.* **171**, 234–244 (2017).
17. T. E. Törnqvist, N. P. Hijma, *Nat. Geosci.* **5**, 601–606 (2012).
18. B. G. Thom, L. D. Wright, J. M. Coleman, *J. Ecol.* **63**, 203–232 (1975).
19. S. A. Woodroffe, B. P. Horton, *J. Asian Earth Sci.* **25**, 29–43 (2005).
20. S. Tanabe, K. Hori, Y. Saito, S. Haruyama, A. Kitamura, *Quat. Sci. Rev.* **22**, 2345–2361 (2003).
21. D. S. Collins *et al.*, *Nat. Commun.* **8**, 15698 (2017).
22. A. Indermühle *et al.*, *Nature* **398**, 121–126 (1999).
23. J. Elsig *et al.*, *Nature* **461**, 507–510 (2009).
24. G. M. Macdonald *et al.*, *Science* **314**, 285–288 (2006).
25. H. J. Poffenbarger, B. A. Needelman, J. P. Megonigal, *Wetlands* **31**, 831–842 (2011).
26. J. Beck *et al.*, *Biogeosciences* **15**, 7155–7175 (2018).
27. T. Sowers, *Quat. Sci. Rev.* **29**, 213–221 (2010).
28. M. Schuerch *et al.*, *Nature* **561**, 231–234 (2018).
29. L. K. Phan, J. S. van Thiel de Vries, M. J. Stive. *J. Coast. Res.* **31**, 233–243 (2014).

ACKNOWLEDGMENTS

We thank J. Mitrovica of Harvard University for providing the GIA model. T. A. Shaw of Nanyang Technology University assisted with the preparation of the figures. Figure 1 used the image library of the Integration and Application Network, University of Maryland Center for Environmental Science (ian.umces.edu/imagelibrary/). **Funding:** N.S. was supported by an Outside Studies Program grant from Macquarie University and AINSE. B.P.H. is supported by the Singapore Ministry of Education Academic Research Fund MOE2018-T2-1-030, the National Research Foundation Singapore, and the Singapore Ministry of Education, under the Research Centers of Excellence initiative. This article is a contribution to PALSEA2 (Palaeo-Constraints on Sea-Level Rise), a working group of the International Union for Quaternary Sciences (INQUA), and International Geoscience Program (IGCP) Project 639, "Sea-Level Changes from Minutes to Millennia." This work is Earth Observatory of Singapore contribution 294. K.R. received funding from the Australian Research Council (FT130100532). **Author contributions:** N.S. conceived the study. N.S., N.S.K., and B.P.H. assembled the contributing mangrove sediment data. E.A. extracted RSLR estimates from GIA models. N.S., E.A., and J.J.K. conducted data analyses. N.S., N.S.K., E.A., and K.R. prepared the figures. All authors contributed to the writing of the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6495/1118/suppl/DC1
Materials and Methods
Figs. S1 to S6
Tables S1 to S2
References (*30*–*124*)
MDAR Reproducibility Checklist
Data S1

18 November 2019; accepted 16 April 2020
10.1126/science.aba2656

IMMUNOLOGY

PIRs mediate innate myeloid cell memory to nonself MHC molecules

Hehua Dai[1]*, Peixiang Lan[2,3]*, Daqiang Zhao[1]*, Khodor Abou-Daya[1], Wentao Liu[2,3], Wenhao Chen[2,3], Andrew J. Friday[1], Amanda L. Williams[1], Tao Sun[4], Jianjiao Chen[4], Wei Chen[4,5], Steven Mortin-Toth[6], Jayne S. Danska[6,7], Chris Wiebe[8], Peter Nickerson[8,9], Tengfang Li[1], Lisa R. Mathews[1], Hêth R. Turnquist[1,10], Matthew L. Nicotra[1,10,11], Sebastien Gingras[10], Eiji Takayama[12], Hiromi Kubagawa[13], Mark J. Shlomchik[10], Martin H. Oberbarnscheidt[1,10,14]†, Xian C. Li[2,3]†, Fadi G. Lakkis[1,10,15]†

Immunological memory specific to previously encountered antigens is a cardinal feature of adaptive lymphoid cells. However, it is unknown whether innate myeloid cells retain memory of prior antigenic stimulation and respond to it more vigorously on subsequent encounters. In this work, we show that murine monocytes and macrophages acquire memory specific to major histocompatibility complex I (MHC-I) antigens, and we identify A-type paired immunoglobulin-like receptors (PIR-As) as the MHC-I receptors necessary for the memory response. We demonstrate that deleting PIR-A in the recipient or blocking PIR-A binding to donor MHC-I molecules blocks memory and attenuates kidney and heart allograft rejection. Thus, innate myeloid cells acquire alloantigen-specific memory that can be targeted to improve transplant outcomes.

Immunological memory protects the host against infection but is also a potent barrier to transplant survival (*1*). Initially thought to be confined to T and B lymphocytes, it is now evident that innate myeloid and lymphoid [natural killer (NK)] cells also acquire features of immunological memory. In lymphoid cells, memory responses are anamnestic (*1*–*3*). By contrast, myeloid cells exhibit memory-like responses, referred to as trained immunity, as secondary stimuli that elicit enhanced reactions need not be related to the primary stimulus (*4*). Whether myeloid cells acquire memory specific to previously encountered antigens is not known but has become an important question, given the key roles of these cells in host defense (*5*).

[1]Thomas E. Starzl Transplantation Institute and Department of Surgery, University of Pittsburgh, Pittsburgh, PA, USA. [2]Immunobiology and Transplant Science Center, Houston Methodist Research Institute, Houston, TX, USA. [3]Department of Surgery, Weill Cornell Medical College of Cornell University, New York, NY, USA. [4]Department of Pediatrics, University of Pittsburgh, Pittsburgh, PA, USA. [5]Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA, USA. [6]Program in Genetics and Genome Biology, Hospital for Sick Children Research Institute, Toronto, ON, Canada. [7]Departments of Immunology and Medical Biophysics, University of Toronto, Toronto, ON, Canada. [8]Department of Medicine, University of Manitoba, Winnipeg, MB, Canada. [9]Department of Immunology, University of Manitoba, Winnipeg, MB, Canada. [10]Department of Immunology, University of Pittsburgh, Pittsburgh, PA, USA. [11]Center for Evolutionary Biology and Medicine (CEBaM), University of Pittsburgh, Pittsburgh, PA, USA. [12]Department of Oral Biochemistry, Asahi University School of Dentistry, Gifu, Japan. [13]Humoral Immune Regulation, Deutsches Rheuma-Forschungszentrum (DRFZ), Berlin, Germany. [14]Center for Critical Care Nephrology, Department of Critical Care Medicine, University of Pittsburgh, Pittsburgh, PA, USA. [15]Department of Medicine, University of Pittsburgh, Pittsburgh, PA, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: mho6@pitt.edu (M.H.O.); xcli@houstonmethodist.org (X.C.L.); lakkisf@upmc.edu (F.G.L.)

We studied monocyte and macrophage reactions to allogeneic bone marrow plug grafts placed under the kidney capsule. Monocytes and macrophages sense major histocompatibility complex (MHC) and non-MHC determinants on allogeneic tissues and mature to antigen-presenting, inflammatory, or cytotoxic cells (*6*–*9*). To investigate memory, B6-*Rag*$^{-/-}$*Il2rg*$^{-/-}$ mice, which lack B, T, NK, and innate lymphoid cells, were immunized with irradiated allogeneic splenocytes. They were then rechallenged with grafts from same-party (i.e., the same source as the immunizing splenocytes) or third-party allogeneic donors after 7, 28, or 49 days. B6-*Rag*$^{-/-}$*Il2rg*$^{-/-}$ (H-2b) mice immunized with BALB/c (H-2d) splenocytes and rechallenged with same-party BALB/c allografts 7 or 28 days later exhibited significantly greater graft infiltration with recipient monocyte-derived dendritic cells (Mo-DCs) than unimmunized recipients or those immunized with syngeneic (B6) or third-party allogeneic (C3H or H-2k) splenocytes (Fig. 1A). This enhanced response dissipated by 49 days after immunization. B6-*Rag*$^{-/-}$*Il2rg*$^{-/-}$ recipients of C3H grafts mounted a heightened response if previously immunized with C3H but not B6 or BALB/c splenocytes, which further demonstrates allospecificity (Fig. 1B). Enhanced responsiveness could not be attributed to immunogen persistence because neither donor splenocytes nor intact donor MHC-I molecules were detected in the host's circulation or spleen at the time of rechallenge (fig. S1A). Thus, host monocytes appear to acquire specific memory to previously encountered alloantigens.

To rule out any possible contribution of lymphoid cells in the graft, we repeated the experiments using *Rag*$^{-/-}$*Il2rg*$^{-/-}$ donors. B6-*Rag*$^{-/-}$*Il2rg*$^{-/-}$ recipients responded more vigorously to BALB/c-*Rag*$^{-/-}$*Il2rg*$^{-/-}$ bone marrow plug allografts if previously immunized with BALB/c-*Rag*$^{-/-}$*Il2rg*$^{-/-}$ but not NOD-*Rag*$^{-/-}$*Il2rg*$^{-/-}$ splenocytes (Fig. 1C), whereas NOD-*Rag*$^{-/-}$*Il2rg*$^{-/-}$ immunization enhanced the response to NOD-*Rag*$^{-/-}$*Il2rg*$^{-/-}$ allografts (Fig. 1D). Thus, monocyte memory to alloantigens is not dependent on lymphoid cells from either the graft or the recipient.

The primary innate alloresponse that generates Mo-DCs is mediated by inflammatory (Ly6C^{hi}) monocytes (*6*, *7*). To test whether this monocyte subset acquires memory, we sorted Ly6C^{hi} monocytes from B6-*Rag*$^{-/-}$*Il2rg*$^{-/-}$ mice 7, 21, or 35 days after immunization and transferred them to naïve B6-*Rag*$^{-/-}$*Il2rg*$^{-/-}$ hosts (Fig. 1E). Mice that received monocytes from allostimulated donors mounted a significantly greater response to BALB/c allografts than those that received monocytes from donors stimulated with syngeneic cells. This demonstrates that Ly6C^{hi} monocytes are capable of mediating memory and that they retain memory function for up to several weeks after their first encounter with an alloantigen.

Macrophages acquire the ability to kill allogeneic targets if first challenged with allogeneic cells along with CD40 receptor cross-linking (*9*). We therefore examined whether macrophages also acquire allospecific memory. B6-*Rag*$^{-/-}$*Il2rg*$^{-/-}$ mice were left unstimulated (naïve) or were immunized with BALB/c splenocytes plus agonistic anti-CD40 antibody, and an in vivo killing assay was performed 14 and 28 days later (Fig. 1F). BALB/c-immunized mice killed BALB/c but not syngeneic (B6) or third-party (C3H) targets, which suggests that macrophages, like monocytes, acquire memory specific to previously encountered alloantigens.

MHC proteins are the principal alloantigens in humans and mice. To test whether monocyte or macrophage memory is dependent on the detection of nonself MHC molecules, we immunized B6-*Rag*$^{-/-}$*Il2rg*$^{-/-}$ (H-2b) recipients with irradiated BALB/c (H-2d) splenocytes and rechallenged them with BALB.B (H-2b) grafts. BALB.B and B6 mice share the same MHC (H-2) but differ at non-MHC loci. Immunization with BALB/c cells did not enhance the response to BALB.B grafts, which indicates that monocyte memory is against nonself MHC molecules (Fig. 1G). The same was true for macrophages (Fig. 1H). Moreover, BALB/c-immunized hosts killed β2-microglobulin knockout targets less efficiently than wild-type cells, which indicates that the response is primarily to nonself MHC-I molecules (Fig. 1H). Furthermore, macrophages from BALB/c-immunized B6 mice bound allogeneic H-2Dd (and, to some extent, allogeneic H-2Kd) but not syngeneic H-2Db and H-2Kb MHC-I tetramers (Fig. 1I). Additionally, host exposure to both MHC and non-MHC polymorphisms at the time of

priming was necessary to generate memory (fig. S1B). Polymorphism in the non-MHC signal regulatory protein alpha (*Sirpa*) gene, which is known to trigger the primary monocyte alloresponse (*8*), was required (fig. S1C).

A database search across 17 mouse genomes for potential MHC-I–binding molecules on myeloid cells identified several families of polymorphic immunoglobulin (Ig) superfamily receptors. Among these were the paired Ig-like receptors (PIRs), which are orthologs of human leukocyte Ig-like receptors (LILRs) (*10*). Closely linked genes coding for six PIR-A proteins and one PIR-B protein are located on mouse chromosome 7 (*11*). PIR-B is nonpolymorphic

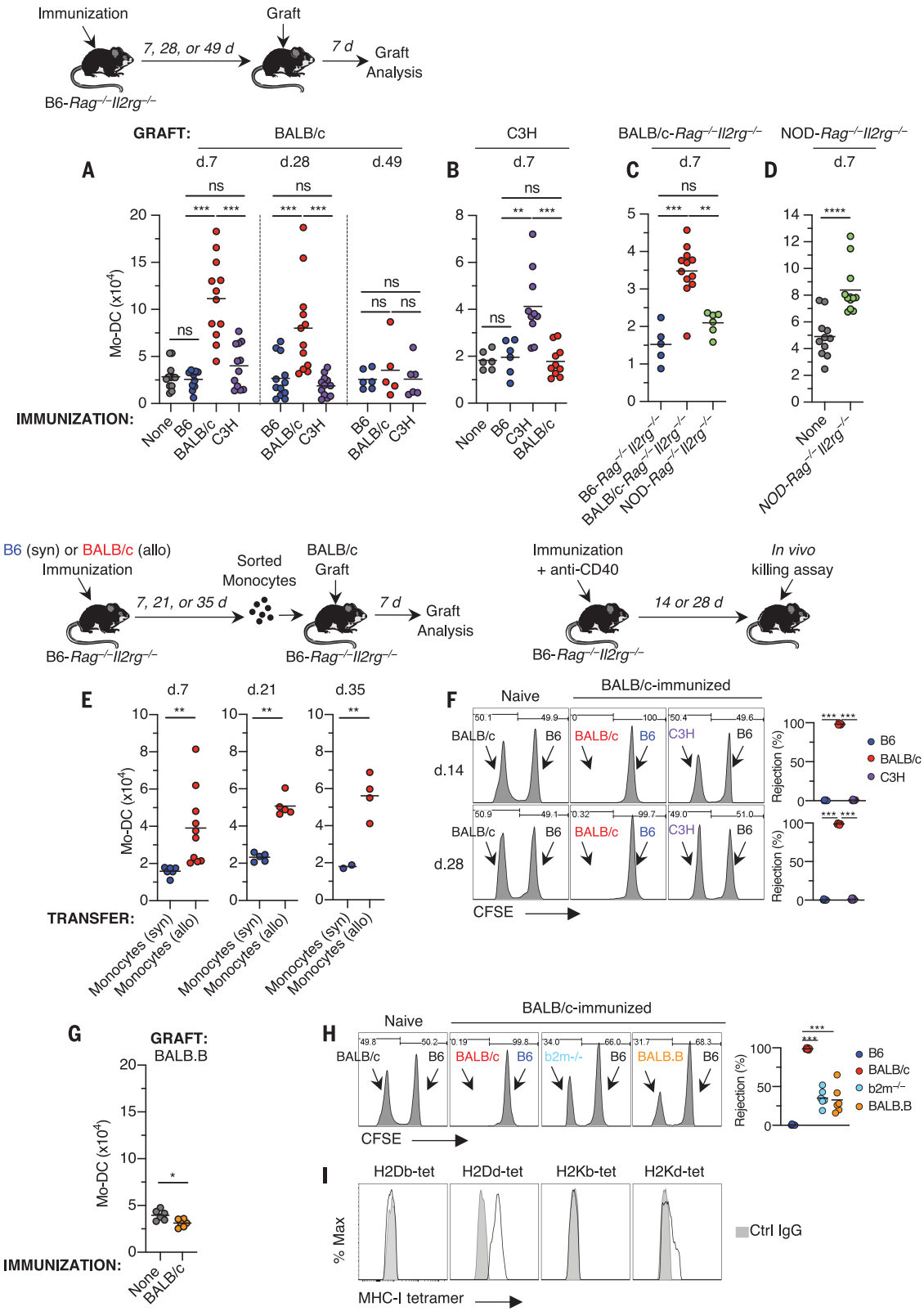


Fig. 1. Monocyte and macrophage memory specific to allogeneic MHC molecules. (**A** to **D**) Monocyte responses in immunized mice measured as monocyte-derived dendritic cells (Mo-DCs) in the graft (n = 5 to 12 mice; N = 1 to 2 experiments). d, day. (**E**) Monocytes from immunized mice transfer memory to naïve recipients (n = 2 to 4; N = 1 to 3). syn, syngeneic; allo, allogeneic. (**F**) Allospecific killing of carboxyfluorescein diacetate succinimidyl ester (CFSE)–labeled cells in immunized mice (n = 6; N = 2). (**G** and **H**) Nonself MHC recognition is necessary for eliciting monocyte (n = 6; N = 1) (G) and macrophage (n = 6; N = 2) (H) memory. Killing of target cells lacking nonself MHC (BALB.B) or MHC-I ($B2m^{-/-}$) is significantly diminished (H). Assays were performed 7 days after immunizing B6-$Rag^{-/-}Il2rg^{-/-}$ mice. (**I**) Allogeneic MHC-I tetramer (tet) binding to macrophages from BALB/c-immunized B6- $Rag^{-/-}Il2rg^{-/-}$ hosts (representative of three experiments). Max, maximum. Statistical analyses: mean and individual biological replicates [(A) to (H)]; one-way analysis of variance (ANOVA) [(A) to (C), (F), and (H)]; or two-tailed Student's t test [(D), (E), and (G)]. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001; ns, not significant.

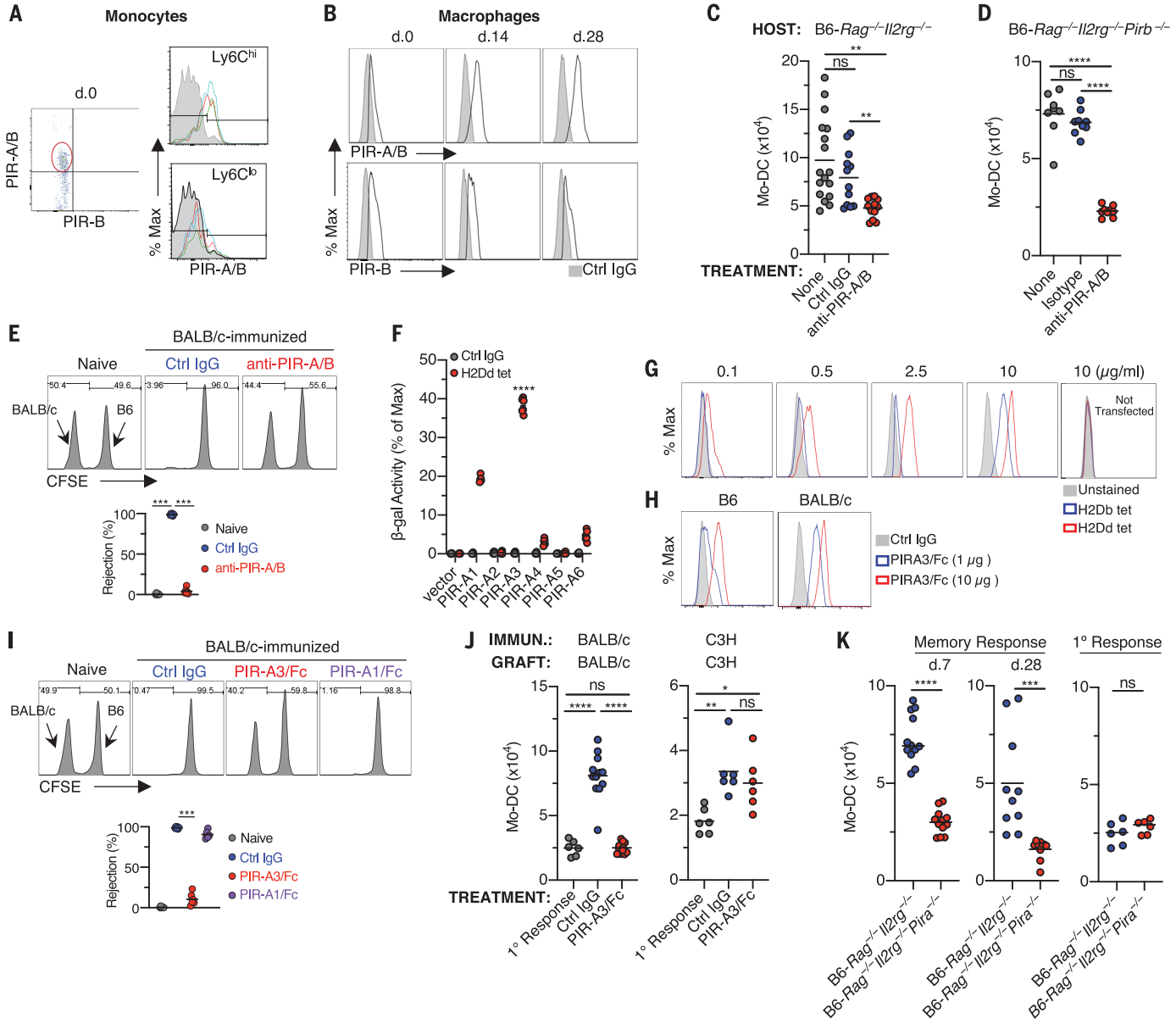


Fig. 2. Memory to allogeneic MHC-I is mediated by PIR-A molecules on monocytes and macrophages. (**A** and **B**) PIR-A and -B expression on Ly6C^{hi} monocytes (A) and macrophages (B) before [day 0 (d.0)] and after alloimmunization (d.14 and d.28). Colored histograms show biological replicates (n = 3). (**C** to **E**) Blocking PIR-A/B inhibits monocyte (n = 12; N = 2) [(C) and (D)] and macrophage (n = 6; N = 2) (E) memory. B6-*Rag*$^{-/-}$*Il2rg*$^{-/-}$ hosts were immunized with BALB/c splenocytes and rechallenged with BALB/c allografts 7 days later. (**F** to **H**) H-2Dd MHC-I tetramers bind preferentially to PIR-A3-transfected BWZ.36 (n = 5; N = 1) (F) and 3T3 cells (G). PIR-A3/Fc binds preferentially to H-2Dd^{+} (BALB/c) cells (H). β-gal; β-galactosidase. (**I** and **J**) Specific inhibition of macrophage (n = 5 to 6; N = 2) (I) and monocyte (n = 6; N = 1 to 2) (J) memory to BALB/c allografts by PIR-A3/Fc. Immun., Immunization. (**K**) Absent memory 7 and 28 days after immunization, but normal primary monocyte alloresponse in *Pira*$^{-/-}$ hosts (n = 6; N = 1 to 2). Statistical analyses: representative of three experiments [(B), (G), and (H)]; mean and individual biological replicates [(C), (D), (F), and (I) to (K)]; one-way ANOVA [(C) to (E), (I), and(J)]; or two-tailed Student's t test (K).

and binds a wide spectrum of MHC-I molecules (*12*). Its cytosolic domain transmits inhibitory signals (*12*, *13*), whereas PIR-As are stimulatory (*14*, *15*). PIR-A isoforms are believed to bind distinct MHC-I molecules (*12*), and PIR-A expression is up-regulated upon myeloid cell differentiation and activation (*15*). Using anti–PIR-A/B and anti–PIR-B antibodies, we detected PIR-A expression on resting Ly6C^{hi} monocytes (Fig. 2A). Macrophages expressed both PIR-A and PIR-B and up-regulated PIR-A after allostimulation (Fig. 2B). Blocking these receptors with an anti–PIR-A/B antibody inhibited monocyte memory in *Pirb*-sufficient and *Pirb*$^{-/-}$ B6-*Rag*$^{-/-}$*Il2rg*$^{-/-}$ hosts (Fig. 2, C and D) and suppressed the killing of allogeneic targets in alloimmunized B6-*Rag*$^{-/-}$*Il2rg*$^{-/-}$ mice (Fig. 2E). We confirmed these results by treating animals with PIR-A3/Fc fusion protein, which preferentially blocks PIR-A3 from binding to its MHC-I ligand, H-2Dd. The preferred ligand of PIR-A3 is H-2Dd (Fig. 2, F to H). H2-Dd tetramers triggered the strongest signal in reporter cells expressing PIR-A3 (Fig. 2F). They bound PIR-A3–transfected 3T3 fibroblasts better than H-2Db tetramers (Fig. 2G), and the binding of PIR-A3/Fc was greater to BALB/c (H-2Dd^{+}) than B6 cells (H-2Db^{+} but H-2Dd^{-}) (Fig. 2H). PIR-A3/Fc inhibited both macrophage (Fig. 2I)

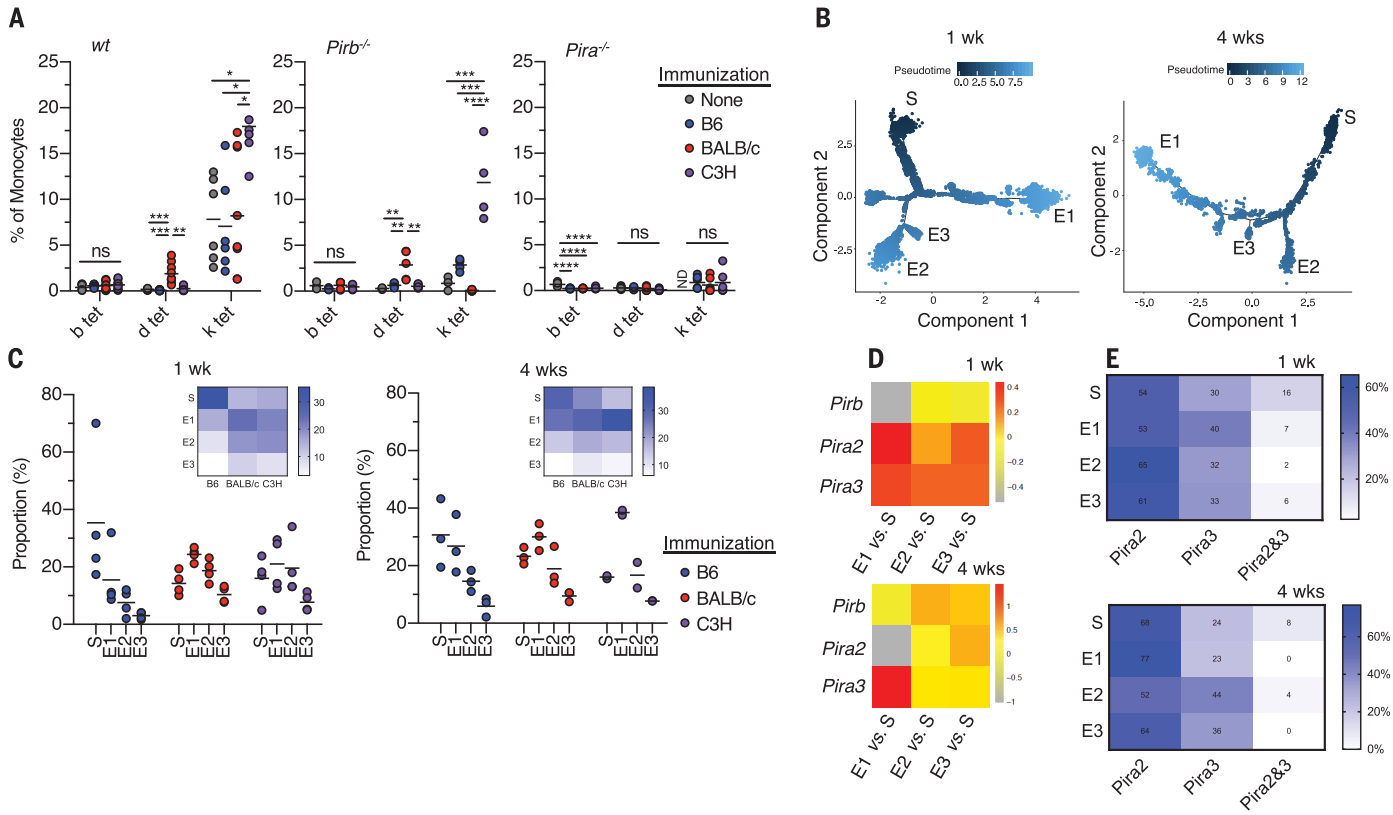


Fig. 3. Mechanisms of monocyte memory. (**A**) Tetramer-positive splenic monocytes 7 days after immunizing B6-*Rag*$^{-/-}$*Il2rg*$^{-/-}$ mice (mean and individual biological replicates; n = 3; N = 1 to 3). wt, wild-type. (**B** to **E**) Splenic monocyte scRNA-seq analysis 1 and 4 weeks after immunization (n = 3; N = 1) illustrated with: pseudotime plots (S indicates starting state, E1 to E3 indicate expanding states) (B); heatmaps and graphs depicting the proportion of S and E states in each immunization group (C); differential expression of *Pir* genes in E states versus the S state (D); and proportion of monocytes expressing *Pira2*, *Pira3*, or both in each state (E). One-way ANOVA (A).

and monocyte (Fig. 2J) memory in BALB/c-immunized B6-*Rag*$^{-/-}$*Il2rg*$^{-/-}$ mice. By contrast, PIR-A1/Fc did not inhibit the killing of BALB/c (H-2d) targets (Fig. 2I). Furthermore, PIR-A3/Fc failed to suppress the monocyte memory response to C3H (H-2k) grafts (Fig. 2J), which confirmed the specificity of PIR-A3 to BALB/c H-2Dd molecules. Finally, we generated PIR-A locus knockout (*Pira*$^{-/-}$) mice lacking PIR-A1-5 (fig. S2), which were bred onto the B6-*Rag*$^{-/-}$*Il2rg*$^{-/-}$ background. Notably, these mice did not mount a monocyte memory response to BALB/c allografts 7 or 28 days after priming (Fig. 2K). By contrast, the primary response to BALB/c allografts (Fig. 2K) and primary dendritic cell and macrophage functions (antigen presentation and IL-6 production) (fig. S3) were not inhibited in *Pira*$^{-/-}$ recipients. Thus, PIR-A mediates innate memory.

We next investigated how monocyte memory is acquired and whether it resembles NK memory (*16*). B6 mice immunized with BALB/c (H-2d) cells selectively increased the proportion and number of splenic monocytes that bind the H-2Dd tetramer, whereas C3H (H-2k) immunization exclusively increased H-2Dk-binding cells (Fig. 3A and fig. S4A). This effect was preserved in *Pirb*$^{-/-}$ mice but abolished in *Pira*$^{-/-}$ mice. Only a minority of monocytes bound more than one type of tetramer (fig. S4B), which suggests that PIR-A expression is variegated. Single-cell RNA sequencing (scRNA-seq) analysis of splenic monocytes performed 1 and 4 weeks after immunization showed that monocytes followed a pseudotime trajectory from a starting state (S), which dominated in mice stimulated with syngeneic cells, to expanding states (E1 to E3), which increased proportionally in allostimulated groups (Fig. 3, B and C). E states were enriched for cell cycle, antigen presentation, and immune pathways such as allograft rejection and graft-versus-host disease (data S1 and S2) and exhibited increased *Pira* and reduced *Pirb* expression (Fig. 4D). Cell cycle pathway enrichment is consistent with prior evidence of monocyte proliferation after allostimulation (*8*). Finally, monocytes that expressed mRNA of two different *Pira* alleles constituted a small minority before and after allostimulation (Fig. 4E), which provides further evidence for variegated PIR-A expression. Thus, a possible mechanism of monocyte memory is the clonal expansion of cells specific to a particular nonself MHC molecule.

Mo-DCs and macrophages contribute to allograft rejection in mice and humans (*7*, *17*–*19*). However, it is unknown whether monocyte or macrophage memory mediated by PIR-A plays a role in this rejection. Kidney allografts survived long-term (>125 days) and maintained normal serum creatinine in lymphoid cell–sufficient wild-type and *Pira*$^{-/-}$ recipients, but they were rapidly rejected by *Pirb*$^{-/-}$ mice (Fig. 4, A and B), consistent with the inhibitory functions of PIR-B (*12*, *13*). PIR-A3/Fc prevented acute allograft rejection in *Pirb*$^{-/-}$ mice (Fig. 4, A and B), which suggests that PIR-A accelerates rejection in the absence of PIR-B. Chronic rejection and cellular infiltrates in wild-type mice were significantly attenuated in *Pira*$^{-/-}$ mice and in *Pirb*$^{-/-}$ recipients treated with PIR-A3/Fc (Fig. 4, C and D, and fig. S5A). As expected, chronic rejection did not occur in grafts that were rejected acutely [*Pirb*$^{-/-}$ and *Pirb*$^{-/-}$ plus control immunoglobulin G (Ctrl IgG) groups] (Fig. 4, C and D). We also observed a pathogenic role for PIR-A in heart transplantation. CTLA4-Ig or PIR-A3/Fc alone caused significant but modest prolongation of allograft survival, whereas their combined administration markedly extended survival and prevented pathology associated with acute or chronic rejection (Fig. 4, E and F, and fig. S5, B and C). Neither PIR-A deficiency nor blockade suppressed alloantibody production (fig. S5, D and E). Thus, PIR-A, which is necessary for monocyte and macrophage memory,

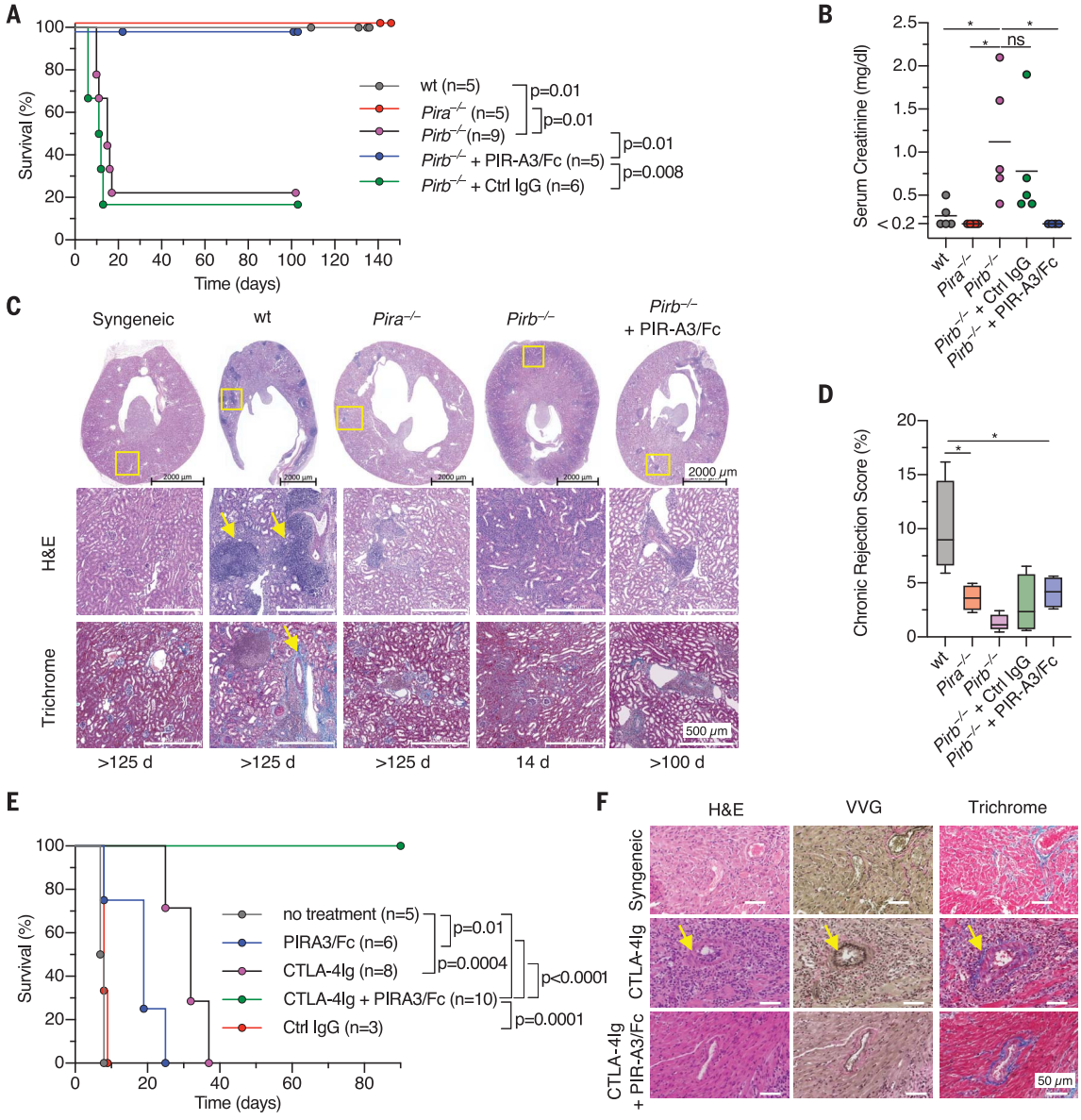


Fig. 4. Genetic deletion or blocking PIR-A attenuates kidney and heart allograft rejection. (**A** to **D**) Survival (A), serum creatinine (B), histology (C), and chronic rejection scores (D) of BALB/c kidneys transplanted to wild-type (wt), *Pira*$^{-/-}$, or *Pirb*$^{-/-}$ B6 recipients. Arrows in (C) point to infiltrates and fibrosis. H&E, hematoxylin and eosin. (**E** and **F**) Survival (E) and histology (F) of BALB/c hearts transplanted to B6 mice treated with PIR-A3/Fc (± CTLA4Ig) or control mouse IgG1 (Ctrl IgG). Arrows in (F) point to chronic rejection vasculopathy. VVG, Verhoeff-Van Gieson. Statistical analyses: mean and individual biological replicates [(B) and (D)]; log-rank [(A) and (E)]; and one-way ANOVA [(B) and (D)].

promotes allograft rejection, whereas PIR-B tempers it.

The findings described in this work identify a pathway of MHC-I allorecognition that is mediated by monocytes and macrophages, generates memory specific to previously encountered MHC-I alloantigens, and contributes to allograft rejection. This pathway extends the domain of classical immunological memory to innate myeloid cells and can be potentially targeted to improve the survival of life-saving organ transplants.

REFERENCES AND NOTES

1. A. Valujskikh, F. G. Lakkis, *Immunol. Rev.* **196**, 65–74 (2003).
2. J. G. O'Leary, M. Goodarzi, D. L. Drayton, U. H. von Andrian, *Nat. Immunol.* **7**, 507–516 (2006).
3. J. C. Sun, J. N. Beilke, L. L. Lanier, *Nature* **457**, 557–561 (2009).
4. M. G. Netea *et al.*, *Science* **352**, aaf1098 (2016).
5. K. Bassler, J. Schulte-Schrepping, S. Warnat-Herresthal, A. C. Aschenbrenner, J. L. Schultze, *Annu. Rev. Immunol.* **37**, 269–293 (2019).
6. D. Zecher, N. van Rooijen, D. M. Rothstein, W. D. Shlomchik, F. G. Lakkis, *J. Immunol.* **183**, 7810–7816 (2009).
7. M. H. Oberbarnscheidt *et al.*, *J. Clin. Invest.* **124**, 3579–3589 (2014).
8. H. Dai *et al.*, *Sci. Immunol.* **2**, eaam6202 (2017).
9. W. Liu, X. Xiao, G. Demirci, J. Madsen, X. C. Li, *J. Immunol.* **188**, 2703–2711 (2012).
10. T. Takai, *Immunology* **115**, 433–440 (2005).
11. H. Kubagawa, P. D. Burrows, M. D. Cooper, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 5261–5266 (1997).
12. A. Nakamura, E. Kobayashi, T. Takai, *Nat. Immunol.* **5**, 623–629 (2004).
13. G. Ma *et al.*, *Immunity* **34**, 385–395 (2011).
14. A. Maeda, M. Kurosaki, T. Kurosaki, *J. Exp. Med.* **188**, 991–995 (1998).
15. H. Kubagawa *et al.*, *J. Exp. Med.* **189**, 309–318 (1999).
16. T. Nabekura, L. L. Lanier, *J. Exp. Med.* **211**, 2455–2465 (2014).
17. Q. Zhuang *et al.*, *Nat. Commun.* **7**, 12623 (2016).
18. M. S. Braza *et al.*, *Immunity* **49**, 819–828.e6 (2018).
19. I. Batal *et al.*, *J. Am. Soc. Nephrol.* **26**, 3102–3113 (2015).

ACKNOWLEDGMENTS

We would like to thank C. Bi and Z. Kou of the Transgenic and Gene Targeting Core (Department of Immunology, University of Pittsburgh School of Medicine) for microinjection of zygotes and production of *Pira*- and *Pirb*-mutant mice. **Funding:** This work was supported by NIH grants AI145881 (M.H.O.), AI080779 (X.C.L.), and AI099465 (F.G.L.) and by funds from the Frank and Athena Sarris Chair in Transplantation Biology at the University of Pittsburgh (F.G.L.). J.S.D. was supported by the JDRF, the Canadian Institutes of Health Research, and the SickKids Foundation. **Author contributions:** H.D., P.L., D.Z., W.L., Wen.C., A.J.F., T.L., L.R.M., and H.R.T. performed experiments and analyzed data. K.A.-D., T.S., and J.C. performed bioinformatic analysis. Wei C. supervised bioinformatic analysis. A.L.W. and S.M.-T. performed mouse breeding and genotyping. J.S.D. supervised generation and genotyping of NOD congenic mice and edited the manuscript. C.W. and P.N. provided data and insights into human correlates of mouse findings. M.L.N. performed comparative mouse genomics and edited the manuscript. S.G. designed and constructed gene knockout mice and contributed to the manuscript writing. E.T. and H.K. designed, constructed, and provided expression vectors and reagents and edited the manuscript. M.J.S. supervised design and construction of *Pirb*$^{-/-}$ mice, provided the mice, and edited the manuscript. M.H.O. participated in experimental design, coordinated experiments, trained and supervised personnel, and wrote the manuscript. M.H.O., X.C.L., and F.G.L. conceived the idea, designed experiments, analyzed data, supervised overall work, and

wrote the manuscript. **Competing interests:** The Hospital for Sick Children receives royalties from Trillium Therapeutics, Inc. for SIRPa-Fc, of which J.S.D. receives a fraction. J.S.D. is a coinventor on a patent related to this work, "Role of SIRP alpha polymorphisms in human acute myeloid leukemia." PCT applications are pending in the United States, Canada, and China and have been awarded in Australia (2015) and Japan (2018). **Data and materials availability:** Raw scRNA-seq data have been deposited into the Gene Expression Omnibus (GEO) database (accession no. GSE147596). All other data needed to evaluate the conclusions of the paper are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6495/1122/suppl/DC1
Materials and Methods
Figs. S1 to S5
Tables S1 to S4
References (*20*–*28*)
Data S1 and S2

View/request a protocol for this paper from *Bio-protocol*.

20 March 2019; resubmitted 2 December 2019
Accepted 10 April 2020
Published online 7 May 2020
10.1126/science.aax4040

CANCER

MTOR signaling orchestrates stress-induced mutagenesis, facilitating adaptive evolution in cancer

Arcadi Cipponi[1,2]*, David L. Goode[3,4], Justin Bedo[5,6,7], Mark J. McCabe[2,8], Marina Pajic[1,2], David R. Croucher[1,2], Alvaro Gonzalez Rajal[1], Simon R. Junankar[1,2], Darren N. Saunders[9], Pavel Lobachevsky[3], Anthony T. Papenfuss[5,6,7,10], Danielle Nessem[1], Max Nobis[1,2], Sean C. Warren[1,2], Paul Timpson[1,2], Mark Cowley[2,8], Ana C. Vargas[11], Min R. Qiu[2,12], Daniele G. Generali[13,14], Shivakumar Keerthikumar[3,4], Uyen Nguyen[1], Niall M. Corcoran[15,16,17], Georgina V. Long[18,19,20,21], Jean-Yves Blay[22,23], David M. Thomas[1,2]*

In microorganisms, evolutionarily conserved mechanisms facilitate adaptation to harsh conditions through stress-induced mutagenesis (SIM). Analogous processes may underpin progression and therapeutic failure in human cancer. We describe SIM in multiple in vitro and in vivo models of human cancers under nongenotoxic drug selection, paradoxically enhancing adaptation at a competing intrinsic fitness cost. A genome-wide approach identified the mechanistic target of rapamycin (MTOR) as a stress-sensing rheostat mediating SIM across multiple cancer types and conditions. These observations are consistent with a two-phase model for drug resistance, in which an initially rapid expansion of genetic diversity is counterbalanced by an intrinsic fitness penalty, subsequently normalizing to complete adaptation under the new conditions. This model suggests synthetic lethal strategies to minimize resistance to anticancer therapy.

Genetic diversity is central to adaptation and evolution in cancer. Mutagenesis in malignancies occurs incrementally (*1*, *2*) or in transient bursts (*3*–*7*), providing increased cell-to-cell variation that facilitates adaptation to selective pressures, including anticancer therapies. Paradoxically, mutagenesis induced by some therapies, such as ionizing radiation, forms a core component of treatment for most cancers. The evolutionary effects of mutagenesis appear to be context dependent; they are potentially harmful under favorable conditions yet beneficial under stressed conditions. Stress-induced mutagenesis (SIM) programs that link environmental conditions to genetic diversity are a common atavistic theme across the phylogenetic tree (*8*) and are well described in prokaryotes (*9*) and lower eukaryotes (*10*–*13*).

Here, we sought evidence for cognate programs in cancer. We first looked for accumulation of DNA damage (*14*) in human cancers exposed to therapies not considered directly genotoxic. Phosphorylated histone H2AX (γ-H2AX) was used as a marker of DNA double-strand breaks (DSBs) in pre- and posttreatment samples taken from diverse cohorts of chemo- and radiotherapy-naïve cancer patients. Increased DNA damage was consistently observed in prostate cancer patients treated with androgen deprivation therapy, in breast cancer patients treated with the aromatase inhibitor letrozole, in melanoma patients treated with the BRAF inhibitors dabrafenib or vemurafenib, and in patients with gastrointestinal stromal tumors treated with the KIT inhibitor imatinib (Fig. 1A and fig. S1). DNA DSBs were also increased in patient-derived xenograft (PDX) pancreatic cancer models treated with the CDK4/6 inhibitor palbociclib and the epidermal growth factor receptor (EGFR) inhibitor erlotinib, suggesting that increased DNA damage is a recurrent feature in human cancers exposed to nongenotoxic therapies.

We interrogated this phenomenon further in vitro using multiple human cancer cell lines exposed to nongenotoxic drug selection. Drugs specific to genetic targets in each cell line were titrated to near-extinction conditions, consistently generating one to 20 resistant colonies per 100,000 cells (Fig. 1B and table S1), which were analyzed for genetic and molecular features, as well as drug sensitivity (fig. S2). All model systems displayed increased DSBs early during evolution (Fig. 1C), which decreased to baseline during subsequent culture (fig. S3). The levels of DSBs were stable in the absence of selection.

To further characterize this phenomenon, we undertook whole-genome sequencing (average read depth 116×) on single cell–derived clonal populations obtained from untreated and early-phase, drug-resistant 94T778 human liposarcoma and SKMEL28 human melanoma lines. Phylogenetic trees were computed on the basis of copy number changes, revealing in both lines distinct clusters of untreated and drug-resistant clones sharing a common trunk. The observed lengths of the trunks (0.414 and 0.539 arbitrary units for 94T778 and SKMEL28 clones, respectively) were significantly longer than the average branch lengths of the respective parental clones (0.118 in 94T778, P = 1.8e-27, and 0.209 in SKMEL28, P = 1.4e-6), consistent with accelerated genomic evolution induced by the therapeutic bottleneck (Fig. 1D and figs. S4 and S5). To formally estimate mutation rates, we expanded clones using an equal number of generations (20 for 94T778; 21 for SKMEL28). Taking into account differential senescence and cell death (figs. S6 and S7, tables S2 and S3, and materials and methods), we quantified the percentage of de novo single nucleotide variants (SNVs) as the number of subclonal SNVs divided by the number of total unique SNVs. Intraclonal diversity, as measured by the number of de novo SNVs, was significantly higher in the resistant populations in both cell lines (Fig. 1E), consistent with higher mutation rates. We sought evidence for resistance-specific mutational signatures but observed no consistent patterns (fig. S8). We also performed targeted sequencing (figs. S9 to S11) on bulk populations, which revealed transiently increased structural and single nucleotide variation early during drug selection.

To further characterize the dynamics of genomic instability, the 94T778 cell line, containing two amplified targetable oncogenes (CDK4 and MDM2), was exposed to palbociclib or to nutlin-3a, an inhibitor of the p53–MDM2 interaction. Cells were transduced at both early and late time points of evolution with a single-copy fluorescent reporter gene (mCherry). In this system, genomic alterations lead to loss of gene expression. In agreement with the targeted sequencing data, drug-resistant cell populations assayed early during evolution demonstrated a significant loss of fluorescence compared with baseline or late time points (fig. S12). To approximate the

[1]The Kinghorn Cancer Centre, Garvan Institute of Medical Research, Darlinghurst, NSW, Australia. [2]St. Vincent's Clinical School, University of New South Wales, Sydney, NSW, Australia. [3]Peter MacCallum Cancer Centre, Melbourne, VIC, Australia. [4]Sir Peter MacCallum Department of Oncology, The University of Melbourne, Melbourne, VIC, Australia. [5]Bioinformatics Division, Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia. [6]Department of Computing and Information Systems, the University of Melbourne, Parkville, VIC, Australia. [7]Peter MacCallum Cancer Centre, Parkville, VIC, Australia. [8]Kinghorn Centre for Clinical Genomics, Garvan Institute of Medical Research, Darlinghurst, NSW, Australia. [9]School of Medical Sciences, University of New South Wales, NSW, Australia. [10]Department of Medical Biology, University of Melbourne, Melbourne, VIC, Australia. [11]Douglass Hanly Moir Pathology, Turramurra, NSW, Australia. [12]Anatomical and Molecular Oncology Pathology, SYDPATH, St. Vincent's Hospital, Darlinghurst, NSW, Australia. [13]Department of Medical, Surgery and Health Sciences, University of Trieste, Trieste, Italy. [14]Breast Cancer Unit and Translational Research Unit, ASST Cremona, Cremona, Italy. [15]Division of Urology, Royal Melbourne Hospital, Parkville, VIC, Australia. [16]Department of Urology, Peninsula Health, Frankston, VIC, Australia. [17]Department of Surgery, University of Melbourne, VIC, Australia. [18]Melanoma Institute Australia, Wollstonecraft, NSW, Australia. [19]The University of Sydney, Sydney, NSW, Australia. [20]Royal North Shore Hospital and Mater Hospital, Sydney, NSW, Australia. [21]Crown Princess Mary Cancer Centre Westmead Hospital, Sydney, NSW, Australia. [22]Centre Leon Berard and Université Claude Bernard Lyon, Lyon, France. [23]UNICANCER, Paris, France.
*Corresponding author. Email: a.cipponi@garvan.org.au (A.C.); d.thomas@garvan.org.au (D.M.T.)

relative contribution of de novo versus preexisting genetic variation to drug resistance, clonal population dynamics were tracked by engineering 94T778 and the human melanoma A375 cell line with a high-complexity DNA barcode library (*15*). Although we observed selection of preexisting genetic variation, the high frequency of individual barcodes (94T778 line average 69.6%, range 42.9 to 93.3%; A375 line average 17.9%, range 3 to 64.7%) is consistent with a contribution of de novo mutagenesis to resistance (fig. S13). This is also suggested by the late (37 weeks, T3) emergence of subclonal *TP53* mutations (C277F, R273L) induced by nutlin-3a, neither of which is visible at weeks 11 (T1) or 21 (T2) (fig. S11).

As noted, mutagenesis may directly entail a fitness penalty driven by the accumulation of deleterious mutations. To study the counterselective effects of SIM, we examined the clonogenic potential of both parental and resistant lines in the absence of drug selection. We observed universally decreased clonogenicity relative to drug-naïve controls (Fig. 1F). Consistent with these effects, colonies obtained from nutlin-3a–resistant 94T778 contained up to 45% senescent cells at 6 weeks, falling to 6% at 25 weeks, compared with 1.5% in untreated controls (fig. S14). Extinction events may explain the loss of some initially resistant clones; for example, the C238F resistance mutation in *TP53* was detected at 11 weeks under nutlin-3a selection but was lost thereafter (fig. S11). These data support a SIM-induced intrinsic fitness penalty early during adaptation.

To identify common genetic pathways mediating SIM in human cancer, we conducted a genome-wide functional screen using the 94T778 cell line. The presence of multiple

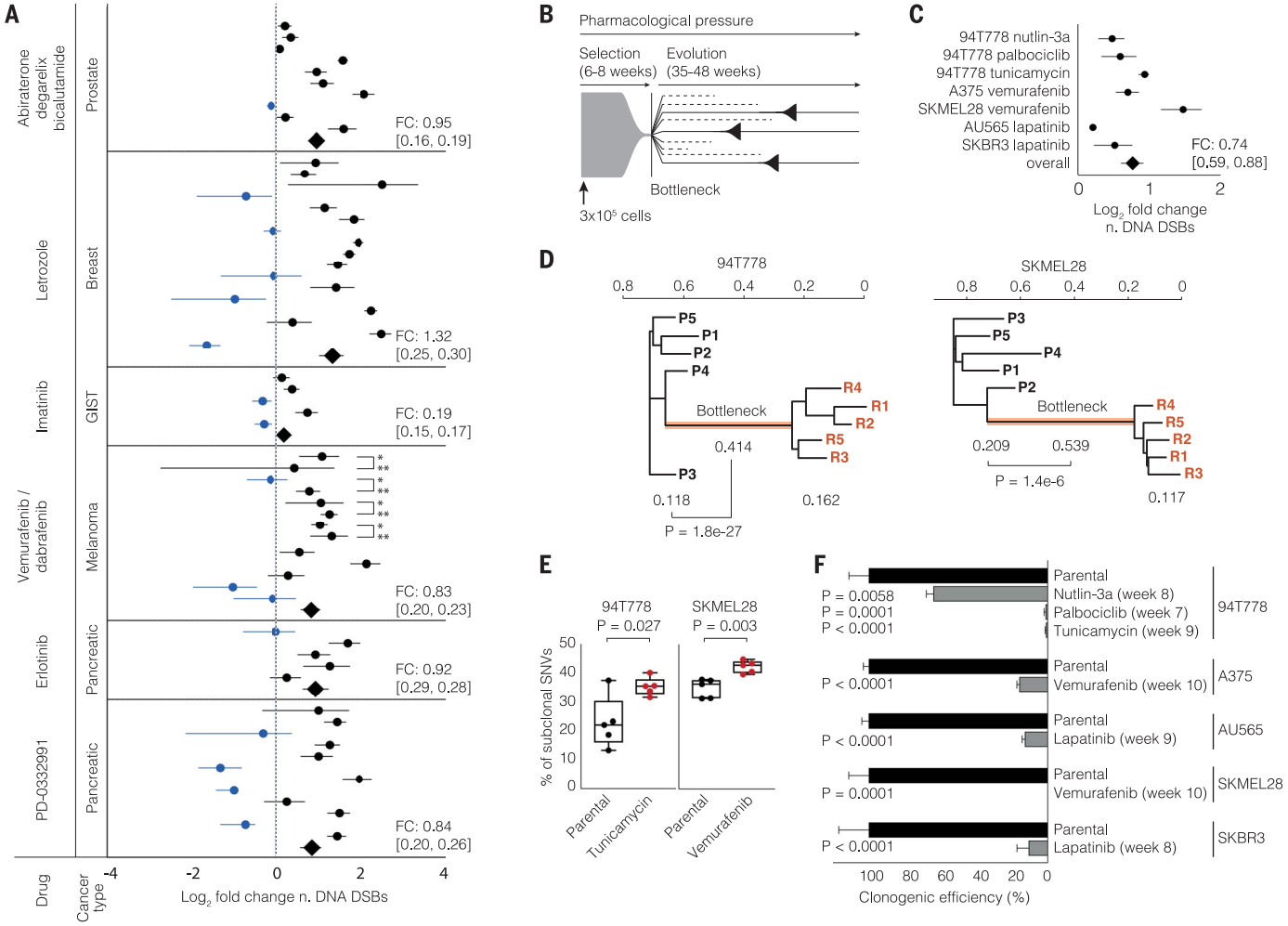


Fig. 1. Selection with targeted anticancer therapies results in genetic diversity. (**A**) Assessment of nuclear DNA DSBs in clinical samples and in PDX pancreatic cancer models. Shown is the fold change (mean ± SD) relative to pretreatment samples in patients and vehicle-treated controls in PDX models. Blue and black symbols represent negative and positive changes, respectively. Black diamonds indicate the overall fold change (FC) with 95% confidence intervals (95% CIs). (**B**) Schematic illustrating the time spans of the selection and evolution phases of in vitro model systems. Shaded area represents total population size, and lines represent the fate of individual clonal lineages after bottleneck. Dashed lines indicate extinction events; solid lines indicate evolution and clonal expansion. (**C**) Fold change values of DNA DSBs relative to untreated parental cells (mean ± SD). (**D**) Phylogenetic analysis of CNVs observed in single cell–derived clonal populations obtained from early-phase, tunicamycin-resistant 94T778 and vemurafenib-resistant SKMEL28 lines. The values below each set of parental (P) and drug-resistant (R) clones represent the average branch length. Values below the common trunks (orange) indicate their length. *P* values were obtained from *Z* tests. (**E**) Quantification of de novo SNVs generated during the expansion of the clonal populations. *P* values were obtained from quasibinomial logistic regression. (**F**) Colony formation assays performed in the absence of drug selection revealing the impact of the adaptive response on cellular fitness during the early phase of evolution. *P* values were obtained from two-tailed *t* test relative to drug-naïve parental cells.

critical drug targets in these cells allowed us to identify shared genetic determinants of resistance. Cells were engineered with a whole-genome short hairpin RNA (shRNA) library and subjected to five different selective conditions (Fig. 2A): (i) culture for 2 weeks without selection to estimate selective drift; (ii) equivalently titrated near-extinction selective conditions using nutlin-3a; (iii) palbociclib; (iv) tunicamycin targeting the unfolded protein response; and (v) incrementally increasing concentrations of nutlin-3a (to determine the relevance of selective stringency).

Initially, the screen identified condition-specific enrichment of shRNAs for genes known to confer resistance, as well as previously unknown genes, in these important pathways ("solution" genes). For examples of known resistance genes, shRNAs targeting *TP53* and *RB1* were among the top-ranked genes for the nutlin-3a and palbociclib conditions, respectively (table S4 and fig. S15). This provided evidence that the assay was performing as expected. We next looked for genes with shRNAs that were enriched across all near-extinction conditions ("facilitator" genes). Facilitator genes were associated with lower levels of enrichment compared with solution genes (table S5), which was perhaps due to an indirect role in mediating resistance through subsequent stochastic events, or possibly to counterselection. Top-ranked among facilitators were the genes encoding the Chaperonin Containing TCP1 Subunit 3 (*CCT3*) and the RNA-processing factor Heterogeneous Nuclear Ribonucleoprotein L (*HNRNPL*) (*16*). CCT3 has an important role in telomere maintenance (*17*, *18*) and genome stability (*19*), and hnRNP protein–mediated alternative splicing has a role in cancer progression (*20*). The next most enriched gene was that coding for the mechanistic target of rapamycin (*MTOR*) (table S6 and fig. S16). In addition, the analysis revealed the enrichment for shRNAs directed to positive regulators of MTOR and the depletion of hairpins targeting the negative regulator *PTEN* (Fig. 2B).

The MTOR signaling pathway is widely expressed in human tissues and functions as an evolutionarily conserved sensor of environmental and endogenous stress (*21*–*23*). The 94T778, SKMEL28, and human breast cancer SKBR3 cell lines were engineered with MTOR-directed shRNAs and were exposed to tunicamycin, vemurafenib, and the EGFR/HER2

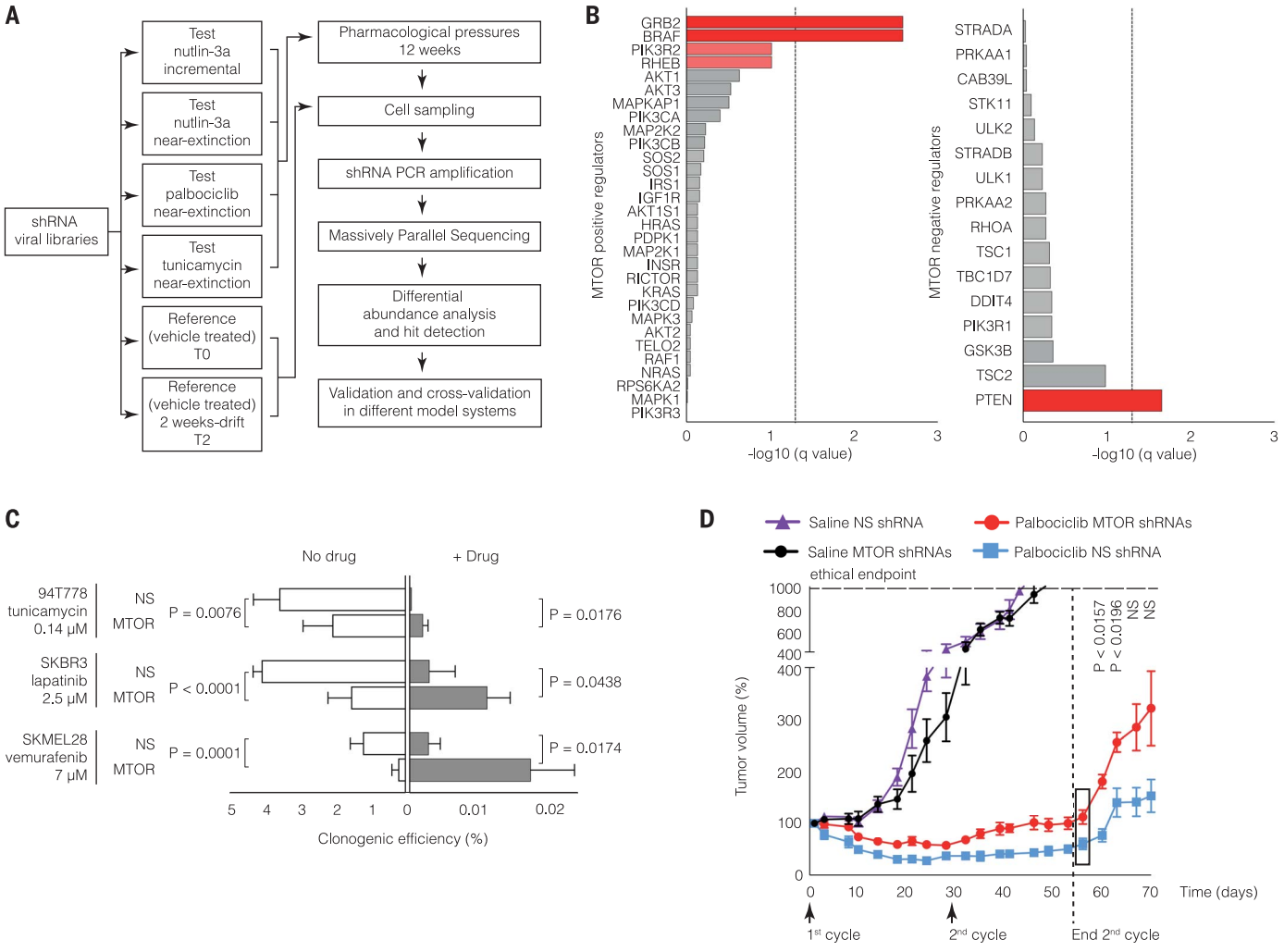


Fig. 2. Whole-genome RNAi screen identifies MTOR as a common evolutionary capacitor in cancer. (**A**) Schematic diagram of the genome-wide RNAi screen. (**B**) Enrichment (left panel) and depletion (right panel) of hairpins targeting components of MTOR signaling. Genes are ranked by *q* values, with genes in red and light red being highly significant and approaching significance, respectively. The dashed line indicates the 0.05 statistical significance threshold. (**C**) Silencing of MTOR signaling imposes a fitness penalty in normal culture conditions (white bars) but fosters adaptability in response to pharmacologic pressures (gray bars). *P* values were obtained from two-tailed *t* test versus nonsilencing (NS) controls. (**D**) Silencing of MTOR affects therapeutic responsiveness to palbociclib in PDX pancreatic cancer models. *P* values were obtained from two-tailed *t* test versus first time point after the end of the second treatment cycle.

inhibitor lapatinib, respectively. All cell lines showed a reduction in MTOR protein levels and a reduced phosphorylation of the downstream kinase p70-S6K (fig. S17A). In addition, we observed a transcriptional repression of genes positively regulated by MTOR and an increased expression of genes repressed by MTOR (*24*) (fig. S17B). As expected, *MTOR* silencing impaired clonogenicity in the absence of selection, but paradoxically increased clonogenicity under drug selection (Fig. 2C and supplementary materials), recapitulating the models described earlier. This effect was not caused by altered drug sensitivity in short-term assays (fig. S18). To exclude off-target effects from gene knock-down, these experiments were validated using shRNAs targeting the 3′ untranslated region of *MTOR*. The stress-induced adaptive potential was abrogated using an RNA interference (RNAi)–resistant open reading frame of the gene (fig. S19). In vivo, repression of MTOR accelerated emergence of palbociclib resistance in a PDX pancreatic cancer model (Fig. 2D) despite enhancing the effect of palbociclib in short-term ex vivo assays (fig. S20). Consistent with previously described kinetics of mutagenesis and impaired intrinsic fitness, a time course analysis of MTOR and p70-S6K phosphorylation levels revealed a significant reduction early during evolution and a restoration to pretreatment levels late during adaptation (fig. S21).

Because inhibition of *MTOR* impairs the DNA-damage response (*25*, *26*), we examined DNA DSBs in 94T778 cells engineered with *MTOR*-specific shRNAs. The analysis of confocal microscopy image data revealed elevated levels of DSBs comparable to those seen in cells spontaneously resistant to nutlin-3a (fig. S22A). Furthermore, the selective MTOR inhibitor torkinib (pp242) delayed the repair of ionizing radiation–induced DNA DSBs, assessed as the number of γ-H2AX (fig. S22B) and 53BP1 (fig. S22C) foci. To quantitate the effects of MTOR silencing on mutagenesis, we undertook whole-genome sequencing on single cell–derived clonal populations obtained from 94T778 cells expressing nonsilencing or MTOR-specific shRNAs. We found a significant increase in intraclonal diversity in MTOR-silenced cells (figs. S23 to S25, tables S2 and S3, and materials and methods). No MTOR-specific mutational signatures were evident in 94T778 clonal populations (fig. S8). In our lines, pharmacologic inhibition of MTOR repressed expression of the homologous recombination (HR) and Fanconi anemia pathways and selectively repressed high-fidelity, but not error-prone DNA polymerases (Fig. 3A and table S7).

To ascertain whether similar transcriptional changes are elicited across a broader range of cancer types, we interrogated the LINCS L1000 database, a transcriptome dataset of 100 lines exposed to ~400 drugs across multiple doses and time points (*27*). Almost identical patterns were seen in multiple cell lines exposed to three different MTOR inhibitors (fig. S26), as well as in 94T778 cells developing spontaneous drug resistance, as described earlier (Fig. 3B and table S8). In the LINCS L1000 database, multiple cytostatic drugs similarly repress accurate DNA repair (fig. S27). This suggests that inhibiting cell proliferation may affect genetic diversity through coordinated regulation of multiple components of DNA repair. To test whether altered DNA repair would enhance adaptation, we reinterrogated the genome-wide screen and found specific enrichment for shRNAs targeting HR genes (table S9). To determine whether the repression of HR directly facilitates drug resistance, we engineered 94T778, SKBR3, and SKMEL28 cells with shRNAs targeting *BRCA1*, *BRCA2*, and *PALB2* and subjected the cells to selection. Recapitulating both spontaneous drug resistance models and silencing of MTOR, impairment of HR reduced clonogenicity in the absence of selection and increased clonogenicity under pharmacologic pressure (Fig. 3C and supplementary materials). The induction of a defective DNA-damage response by non-genotoxic cytostatic therapies suggests synthetic lethality-based strategies. In a preliminary proof-of-concept experiment, we assessed the impact of palbociclib combined with a poly(ADP-ribose) polymerase inhibitor (rucaparib) in a

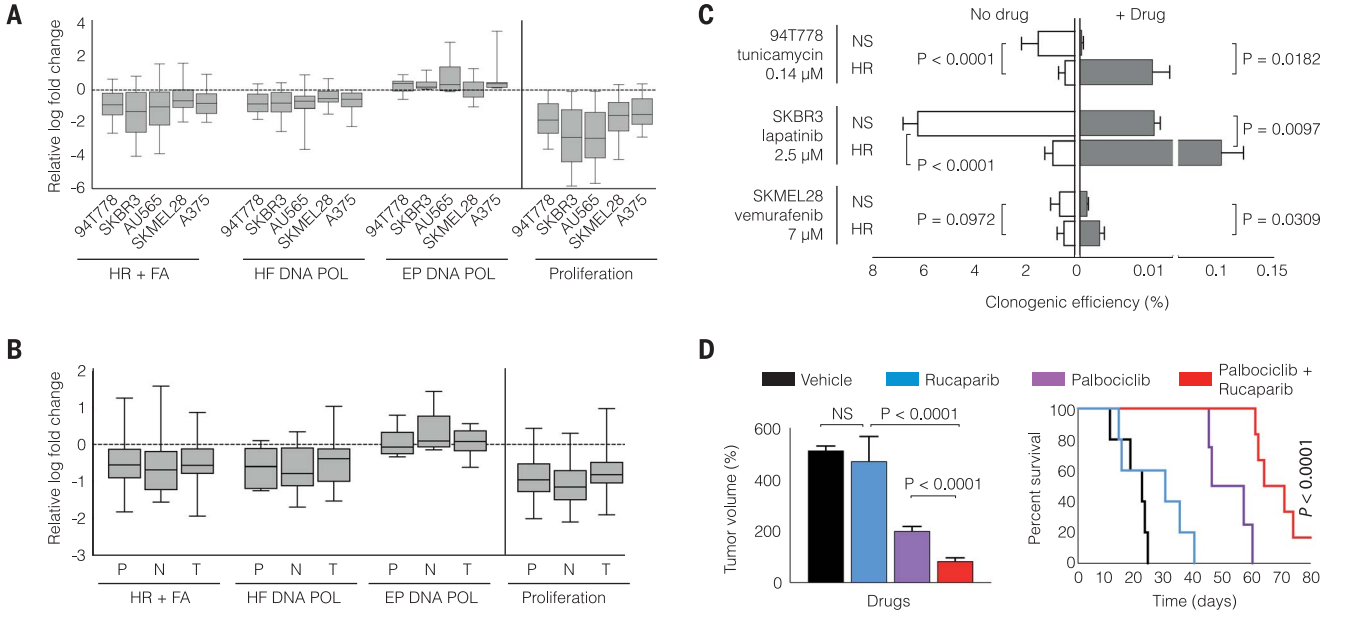


Fig. 3. Inhibition of MTOR fosters adaptive mutagenesis by repressing accurate DNA repair. (**A**) Transcriptional changes induced by torkinib (pp242). *P* values were obtained from one-sided Wilcoxon rank-sum test and are shown in table S7. (**B**) Transcriptional changes observed during the early phase of the adaptive evolution in 94T778 cells exposed to palbociclib (P), nutlin-3a (N), and tunicamycin (T) sampled at 10 weeks for P and N and at 9 weeks after bottleneck for the T condition. *P* values were obtained from one-sided Wilcoxon rank-sum test and are shown in table S8. (**C**) Effect of silencing of HR genes in the absence (white bars) or presence of pharmacologic pressures (gray bars). *P* values were obtained from two-tailed *t* test versus NS control shRNA. (**D**) Effect of palbociclib and rucaparib combination therapy on tumor growth inhibition (bar plot, 30 days) and overall survival (Kaplan–Meier plot) in a PDX pancreatic cancer model. *P* values for tumor growth were obtained from two-tailed *t* test; a log-rank test on palbociclib + rucaparib versus palbociclib monotherapy was used in the Kaplan–Meier plot.

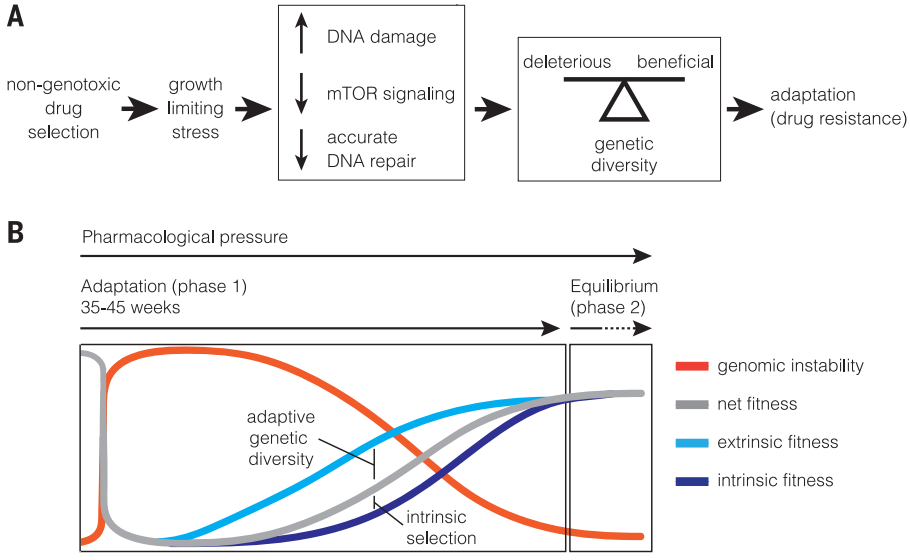


Fig. 4. Conserved mechanisms underpinning SIM in human cancer. (**A**) Selection by cytostatic anticancer therapies leads to increased levels of DNA damage and impaired mTOR signaling, providing genetic diversity and accelerating adaptation to anticancer treatments. (**B**) During the initial phase of adaptation, the fitness landscape is a dynamic balance between the deleterious effect of intrinsic selection and the beneficial effect conferred by resistant genomic configurations. Fit genotypes are progressively stabilized during the second equilibrium phase.

pancreatic model. Combination therapy enhanced the antitumor effects compared with either agent alone. (Fig. 3D).

Collectively, these data reveal SIM as a common mechanism facilitating the evolution of human cancer and are broadly in agreement with recent observations in colorectal cancer (*28*). As with error-prone and faithful DNA polymerases, the advantage conferred by SIM in a metazoan context is not clear and may simply be an evolutionary remnant. MTOR-mediated SIM slows replication and fosters genomic instability by impairing accurate DNA repair, thereby enhancing genetic diversity and facilitating resistance to therapy (*29*) (Fig. 4A). Both suppression of MTOR signaling and DNA repair also directly confer an intrinsic fitness penalty by inhibition of growth signaling and by increasing the accumulation of deleterious mutations, respectively. Rather than being a simple process, our data suggest that drug resistance may be the compound result of two independent but related mechanisms. Initially, net fitness balances the intrinsic fitness penalty of MTOR-mediated SIM with the generation of resistant genotypes undergoing extreme selection. Subsequently, normalization of SIM and fixation of stably resistant genomic configurations establish a new adaptive equilibrium (Fig. 4B). From a clinical perspective, our findings may explain the observation that agents targeting MTOR or the upstream phosphoinositide 3-kinase pathway have greater effects on objective responses or progression-free survival than on overall survival (*30*). From a drug development perspective, enhanced adaptation leading to drug resistance is undetectable by most preclinical assays of anticancer activity or by short-term measures of objective responses in trials. Our findings also provide a rational framework for synthetic lethal combinations of cytostatic agents with genotoxic therapies. Such combinations could potentially generate a lethal mutational load during the initial phase of adaptive evolution, thereby reducing therapeutic failure.

REFERENCES AND NOTES

1. E. R. Fearon, B. Vogelstein, *Cell* **61**, 759–767 (1990).
2. S. Jones *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4283–4288 (2008).
3. P. J. Stephens *et al.*, *Cell* **144**, 27–40 (2011).
4. A. Sottoriva *et al.*, *Nat. Genet.* **47**, 209–216 (2015).
5. R. Gao *et al.*, *Nat. Genet.* **48**, 1119–1130 (2016).
6. F. Notta *et al.*, *Nature* **538**, 378–382 (2016).
7. M. Petljak *et al.*, *Cell* **176**, 1282–1294.e20 (2019).
8. R. S. Galhardo, P. J. Hastings, S. M. Rosenberg, *Crit. Rev. Biochem. Mol. Biol.* **42**, 399–435 (2007).
9. S. M. Rosenberg, C. Shee, R. L. Frisch, P. J. Hastings, *BioEssays* **34**, 885–892 (2012).
10. E. Shor, C. A. Fox, J. R. Broach, *PLOS Genet.* **9**, e1003680 (2013).
11. S. M. Rosenberg, C. Queitsch, *Science* **343**, 1088–1089 (2014).
12. G. Chen, W. D. Bradford, C. W. Seidel, R. Li, *Nature* **482**, 246–250 (2012).
13. B. D. Harrison *et al.*, *PLOS Biol.* **12**, e1001815 (2014).
14. W. M. Bonner *et al.*, *Nat. Rev. Cancer* **8**, 957–967 (2008).
15. H. E. Bhang *et al.*, *Nat. Med.* **21**, 440–448 (2015).
16. L. H. Hung *et al.*, *RNA* **14**, 284–296 (2008).
17. Y. Zhang *et al.*, *Cancer Lett.* **372**, 101–109 (2016).
18. H. Yébenes, P. Mesa, I. G. Muñoz, G. Montoya, J. M. Valpuesta, *Trends Biochem. Sci.* **36**, 424–432 (2011).
19. D. W. Garsed *et al.*, *Cancer Cell* **26**, 653–667 (2014).
20. E. El Marabti, I. Younis, *Front. Mol. Biosci.* **5**, 80 (2018).
21. R. A. Saxton, D. M. Sabatini, *Cell* **169**, 361–371 (2017).
22. J. Aramburu, M. C. Ortells, S. Tejedor, M. Buxadé, C. López-Rodríguez, *Sci. Signal.* **7**, re2 (2014).
23. A. M. Heberle *et al.*, *Mol. Cell. Oncol.* **2**, e970489 (2014).
24. C. J. Creighton *et al.*, *Breast Cancer Res.* **12**, R40 (2010).
25. F. Guo, *J. Oncobiomarkers* **2**, 5 (2014).
26. W. P. Roos, A. D. Thomas, B. Kaina, *Nat. Rev. Cancer* **16**, 20–33 (2016).
27. Q. Duan *et al.*, *Nucleic Acids Res.* **42** (W1), W449-60 (2014).
28. M. Russo *et al.*, *Science* **366**, 1473–1480 (2019).
29. S. L. Carter, A. C. Eklund, I. S. Kohane, L. N. Harris, Z. Szallasi, *Nat. Genet.* **38**, 1043–1048 (2006).
30. F. Janku, T. A. Yap, F. Meric-Bernstam, *Nat. Rev. Clin. Oncol.* **15**, 273–291 (2018).

ACKNOWLEDGMENTS

We thank K. Simpson, P. Madhamshettiwar, and J. Luu for providing the shRNA library and for QC analysis; G. Mir Arnau for assistance with RNA and DNA sequencing; C.L. Chan for assistance with the preparation of DNA libraries; W. Hughes for assistance with confocal microscopy; O. Martin and H. Reza Bigdeli for assistance with quantification of DNA damage; L. Lara-Gonzalez, N. Thio, and M. Zethoven for assistance with bioinformatics; and L. Caldon, G. Rancati, and D. Bowtell for their critical reading of the manuscript. **Funding:** This research was supported by the Australian National Health and Medical Research Council (NHMRC project grant no. 1088353 to D.M.T., A.C., and D.L.G.) and by the Girgensohn Foundation (A.C.). In vivo studies were supported by NHMRC 1162860, NHMRC 1162556, and Cancer Australia 1143699 project grants (M.P.). J.B. was supported by the Stafford Fox Centenary Fellowship in Bioinformatics and Computational Biology of Rare Cancers. **Author contributions:** A.C. and D.M.T. conceived and supervised the study, wrote the manuscript, and acquired research funding; A.C., M.J.M., A.T.P., S.R.J., U.N., and A.G.R. performed the experiments and analyzed the data; D.L.G., S.K., and J.B. performed bioinformatics analyses. D.G.G, N.M.C., G.V.L., M.P, and J.-Y.B. provided the clinical samples; A.C.V. and M.R.Q. analyzed the clinical samples; P.L. developed the JCountPro software. **Competing interests:** D.M.T. is a paid consultant and received research support from Amgen, Eisai, Pfizer, Roche, Astra Zeneca, Novartis, and Bayer. G.V.L. is a paid consultant for Aduro Biotech Inc., Pierre-Fabre Medicament, Bristol-Myers Squibb, Amgen Inc., Merck Sharp & Dohme, Novartis, Array Biopharma Pty Ltd., and Sandos. The other authors declare no competing interests associated with this work. **Data and materials availability:** The plasmids pLV-mCherry, psPAX2, pCAG-VSVG, and pPB-DEST-53BP1$_{trunc}$, are available from AddGene under a material transfer agreement. The pCMV-hyPBase plasmid is available upon request to the Sanger Institute Archives (http://www.sanger.ac.uk/technology/clonerequests/). DNA-sequencing data have been deposited in the NCBI's BioProject database (ID: PRJNA623123). RNA-sequencing data have been deposited in NCBI's Gene Expression Omnibus (accession nos. GSE148342 and GSE148344). Computer codes are available at GitHub (https://github.com/dlgoode/Cipponi_Science_2020).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6495/1127/suppl/DC1
Materials and Methods
Figs. S1 to S27
Tables S1 to S9
References (*31*–*72*)

View/request a protocol for this paper from *Bio-protocol*.

11 September 2018; resubmitted 9 November 2019
Accepted 10 April 2020
10.1126/science.aau8768

REPRODUCTIVE BIOLOGY

NELL2-mediated lumicrine signaling through OVCH2 is required for male fertility

Daiji Kiyozumi[1,2], Taichi Noda[1,2], Ryo Yamaguchi[2,3], Tomohiro Tobita[2,4], Takafumi Matsumura[2,3], Kentaro Shimada[2,3], Mayo Kodani[2,3], Takashi Kohda[5], Yoshitaka Fujihara[1,2]*, Manabu Ozawa[6], Zhifeng Yu[7], Gabriella Miklossy[7], Kurt M. Bohren[7], Masato Horie[8], Masaru Okabe[1,2,3], Martin M. Matzuk[7,9,10,11]†, Masahito Ikawa[1,2,3,4,6]†

The lumicrine system is a postulated signaling system in which testis-derived (upstream) secreted factors enter the male reproductive tract to regulate epididymal (downstream) pathways required for sperm maturation. Until now, no lumicrine factors have been identified. We demonstrate that a testicular germ-cell–secreted epidermal growth factor–like protein, neural epidermal growth factor–like–like 2 (NELL2), specifically binds to an orphan receptor tyrosine kinase, c-ros oncogene 1 (ROS1), and mediates the differentiation of the initial segment (IS) of the caput epididymis. Male mice in which *Nell2* had been knocked out were infertile. The IS-specific secreted proteases, ovochymase 2 (OVCH2) and A disintegrin and metallopeptidase 28 (ADAM28), were expressed upon IS maturation, and OVCH2 was required for processing of the sperm surface protein ADAM3, which is required for sperm fertilizing ability. This work identifies a lumicrine system essential for testis-epididymis-spermatozoa (NELL2-ROS1-OVCH2-ADAM3) signaling and male fertility.

Sex hormones, especially androgens, play critical roles in male genital tract development and function. In addition to androgens, it is hypothesized that proteins and/or small molecules produced in the mammalian testis (upstream) are secreted into the luminal space and are transported to the epididymis to regulate, maintain, and/or differentiate the epididymis (downstream) (*1*). However, the molecular entities of this lumicrine system have not been identified. An orphan receptor tyrosine kinase—c-ros oncogene 1 (ROS1), which is well known as oncogenic when its gene is rearranged to form active fusion proteins—is expressed in the initial segment (IS) of the caput epididymis and is indispensable for postnatal IS differentiation (*2*). Neither the testicular factors that regulate IS differentiation nor the mechanisms by which differentiated IS matures spermatozoa has been fully clarified. In this study, we identified a testis-secreted factor that binds to ROS1 and is indispensable for IS differentiation and male fertility.

The dependency of IS on testicular lumicrine factors is reported to be established between postnatal day 15 (P15) and P19 (*3*). We found that germ cell–deficient *Kit*$^{W/Wv}$ (*W*/*W*^{v}) mice also show impaired IS differentiation (fig. S1). These observations suggest that the putative ROS1 ligand is expressed by developing testicular germ cells (TGCs). To identify the testicular ligand for epididymal orphan receptor ROS1, we screened for testis-secreted proteins that are up-regulated during spermatogenesis (fig. S2). First, from GSE640 Microarray data (36,939 spots), we selected 1706 spots as matrisome genes in silico (*4*). Second, we chose nine candidate proteins that satisfied the criteria that we set (proteins with N-terminal signal sequence, no transmembrane domain, and average transcript levels more than three times higher after P18 compared with those before P14). We excluded seven genes in which knockout (KO) mice lacking these proteins were fertile (*Comp*, *Vit*, *Zp3r*, *C1qtnf4*, and *Agt*) (*5*–*9*) or showed different phenotypes from *Ros1* KO mice (*Insl6* and *Prss21*) (*10*, *11*). Because PRSS39 is a protease and less likely to be a ROS1 ligand, we focused on neural epidermal growth factor–like–like 2 (NELL2), which satisfied the criteria above and could be a hypothetical lumicrine factor and regulator of male fertility.

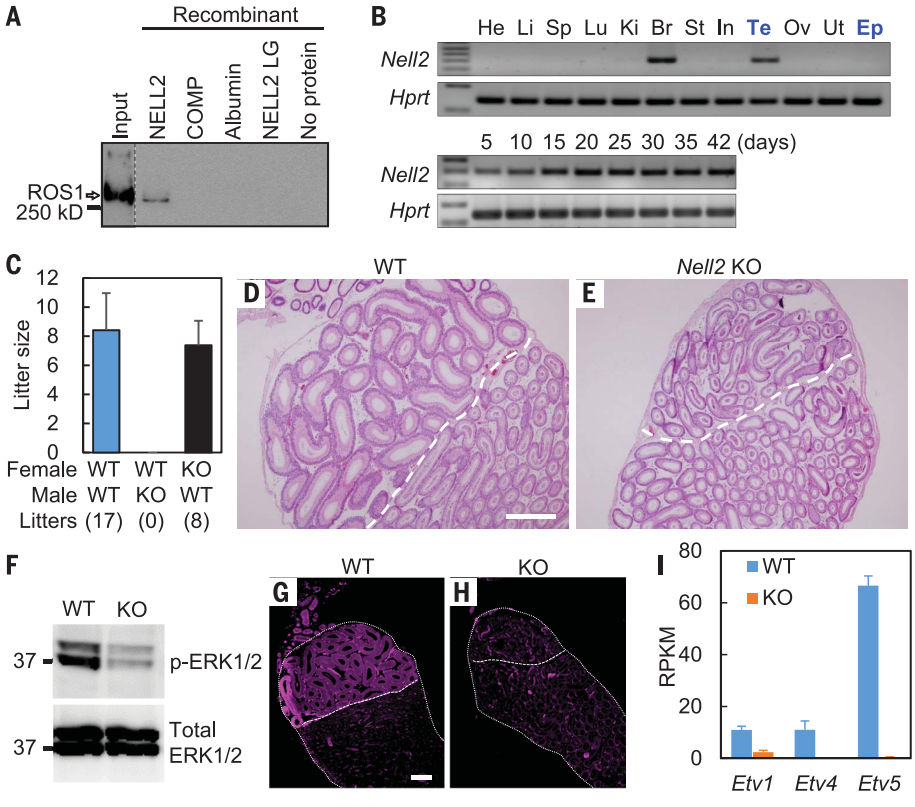


Fig. 1. Testicular NELL2 is indispensable for ROS1-mediated differentiation of epididymal IS. (**A**) In vitro binding of ROS1 with NELL2 and several other indicated proteins. (**B**) Reverse transcription polymerase chain reaction analyses of *Nell2* expression in adult organs and in postnatal testis. *Hprt* is also shown as an internal control. (**C**) Litter sizes of *Nell2* KO mice. Average and SD are shown. (**D** and **E**) Hematoxylin and eosin (H&E) staining of caput epididymis of 14-week-old (D) WT and (E) *Nell2* KO mice. Scale bar, 50 μm. (**F** to **H**) Phospho-ERK immunoblot (F) and immunofluorescence analyses of (G) WT and (H) *Nell2* KO 14-week-old caput epididymis. Scale bar, 50 μm. (**I**) Expression of transcription factors downstream of ERK signaling in 14-week-old WT and *Nell2* KO caput epididymis. RPKM, reads per kilobase per million.

Recombinant NELL2 protein specifically bound recombinant ROS1 in vitro, whereas recombinant cartilage oligomeric matrix protein (COMP), which carries epidermal growth factor–like domains, did not (Fig. 1A and fig. S2). In wild-type (WT) mice, *Nell2* is expressed in testis and brain but not in epididymis (Fig. 1B and fig. S1). In W/W^{v} mice, *Nell2* is expressed in brain, but the IS does not differentiate (fig. S1). Therefore, we focused on testis as the source of NELL2 and identified spermatocytes as *Nell2*-expressing cells with single-cell RNA-sequencing (RNA-seq) data (fig. S1) (*12*), which is consistent with increased *Nell2* transcript levels from P14 (fig. S2). Because spermatocytes are located inside the blood-testis barrier in seminiferous tubules, these data suggest that NELL2 is secreted from TGCs, transits through the luminal space of seminiferous tubules, and binds to ROS1 on the epididymal epithelial cells.

To investigate NELL2 functions in vivo, we generated mice in which *Nell2* had been knocked out (*Nell2* KO mice) using CRISPR-Cas9 (fig. S3) and found that *Nell2* KO males are sterile (Fig. 1C). A *Clgn-Nell2* transgene, which expresses *Nell2* in a TGC-specific manner (*13*), completely rescued the male infertility of the *Nell2* KO mice, again excluding a possibility that NELL2 expressed in the brain regulates fertility (fig. S4). We detected testis-secreted His-tagged NELL2 in the luminal contents of the epididymal duct, supporting that NELL2 is a lumicrine factor (fig. S4).

To determine the cause of infertility in *Nell2* KO males, we examined their reproductive tracts. Spermatogenesis did not appear to be affected in the *Nell2* KO testis (fig. S3). By contrast, the IS was poorly differentiated in *Nell2* KO males (Fig. 1, D and E, and figs. S3 and S5), similar to *Ros1* KO mice (*2*). IS differentiation begins at 2 to 3 weeks of age and corresponds with spatiotemporal *Nell2* expression in testis. Postnatal differentiation of the IS was completely abolished in *Nell2* KO males and continued throughout life (fig. S5).

Extracellular signal–regulated kinase 1/2 (ERK1/2), a signal mediator whose phosphorylation depends on testicular lumicrine factors (*14*), is up-regulated from 2 to 3 weeks postnatally in parallel with IS differentiation (*15*). Whereas the ERK1/2 signal is down-regulated in *Ros1* mutant epididymis (*15*), it is hypophosphorylated in *Nell2* KO caput epididymis (Fig. 1, F to H). Transcription factors *Etv1*, *Etv4*, and *Etv5*, which are located downstream of ERK signaling, are down-regulated in *Nell2* KO IS (Fig. 1I). These results suggest that the molecular basis of impaired IS differentiation is common between *Nell2* KO mice and *Ros1* KO mice.

We next characterized the relationship between impaired IS differentiation and male infertility. When spermatozoa ejaculated into female reproductive tract were visualized with an *Acr-egfp* transgene (*16*, *17*), we observed *Nell2* KO spermatozoa in the uterus but not in the oviduct (Fig. 2, A to F), indicating that the *Nell2* KO spermatozoa are deficient in the passage through the uterotubal junction (UTJ), as observed in *Ros1* KO mice (*18*). Further, both KO mice showed lack of sperm-binding ability to zona pellucida (ZP), the oocyte extracellular matrix (Fig. 2, G and H, and fig. S6).

The inability of spermatozoa to pass through the UTJ and to bind to the ZP is often caused by loss of mature A disintegrin and metallopeptidase 3 (ADAM3), a spermatozoa surface

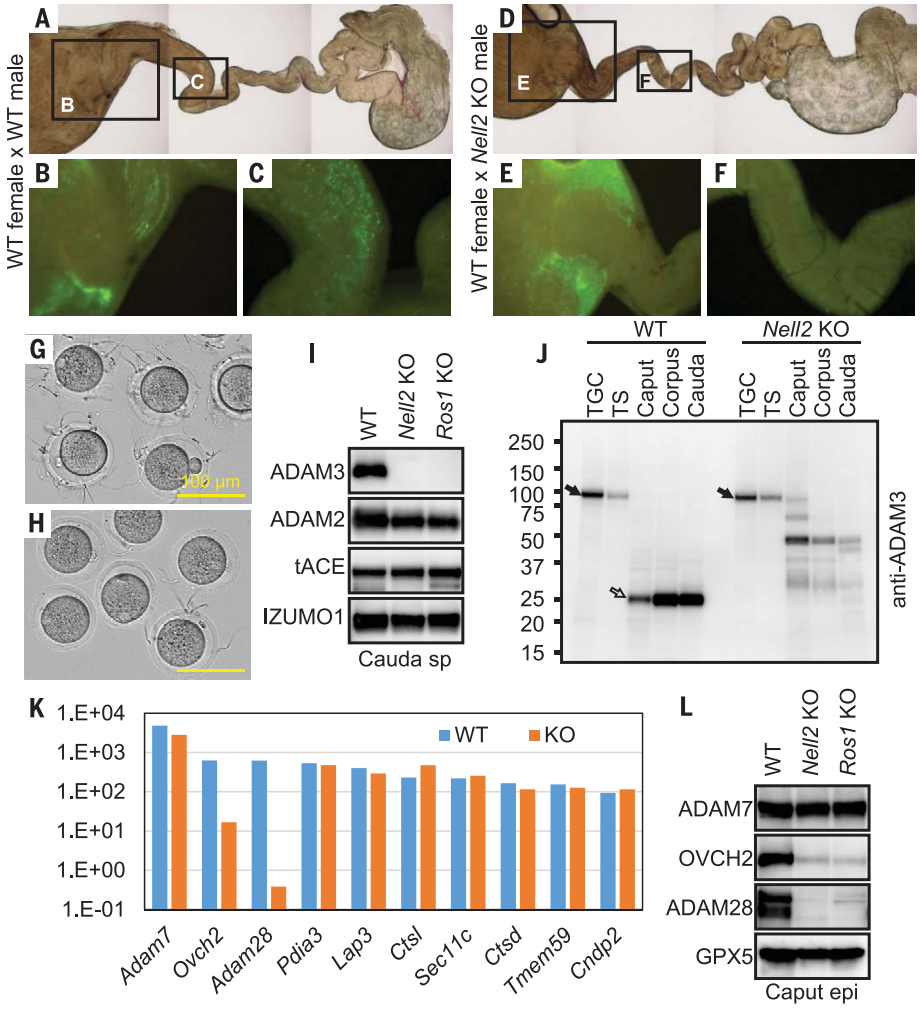


Fig. 2. NELL2 is required for ADAM3 maturation and subsequent sperm fertilizing ability. (**A** to **F**) Migration of [(A) to (C)] green fluorescent protein–tagged WT and [(D) to (F)] *Nell2* KO sperm ejaculated into WT female reproductive tract. (**G** and **H**) Sperm-ZP binding assay with (G) WT and (H) *Nell2* KO spermatozoa. Scale bars, 100 μm. (**I**) Immunoblot detection of cauda epididymal sperm lysates. (**J**) Immunoblot detection of ADAM3 in TGC; TS (testicular spermatozoa); and caput, corpus, and cauda epididymal spermatozoa. (**K**) Expression of genes encoding proteases in WT and *Nell2* KO caput epididymis. Only genes whose average read counts in WT are >100 are shown. (**L**) Immunoblot analysis of caput epididymal protein lysates from WT, *Nell2* KO, and *Ros1* KO males.

[1]Immunology Frontier Research Center, Osaka University, Suita, Osaka 5650871, Japan. [2]Research Institute for Microbial Diseases, Osaka University, Suita, Osaka 5650871, Japan. [3]Graduate School of Pharmaceutical Sciences, Osaka University, Suita, Osaka 5650871, Japan. [4]Graduate School of Medicine, Osaka University, Suita, Osaka 5650871, Japan. [5]Faculty of Life and Environmental Sciences, University of Yamanashi, Kofu, Yamanashi 4008510, Japan. [6]The Institute of Medical Science, The University of Tokyo, Minato-ku, Tokyo 1088639, Japan. [7]Center for Drug Discovery and Department of Pathology and Immunology, Baylor College of Medicine, Houston, TX 77030, USA. [8]Department of CNS Research, Otsuka Pharmaceutical, Kawauchi-cho, Tokushima 771-0192, Japan. [9]Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX 77030, USA. [10]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA. [11]Department of Pharmacology and Chemical Biology, Baylor College of Medicine, Houston, TX 77030, USA.
*Present address: Department of Bioscience and Genetics, National Cerebral and Cardiovascular Center, Suita, Osaka 5648565, Japan.
†Corresponding author. Email: mmatzuk@bcm.edu (M.M.M.); ikawa@biken.osaka-u.ac.jp (M.I.)

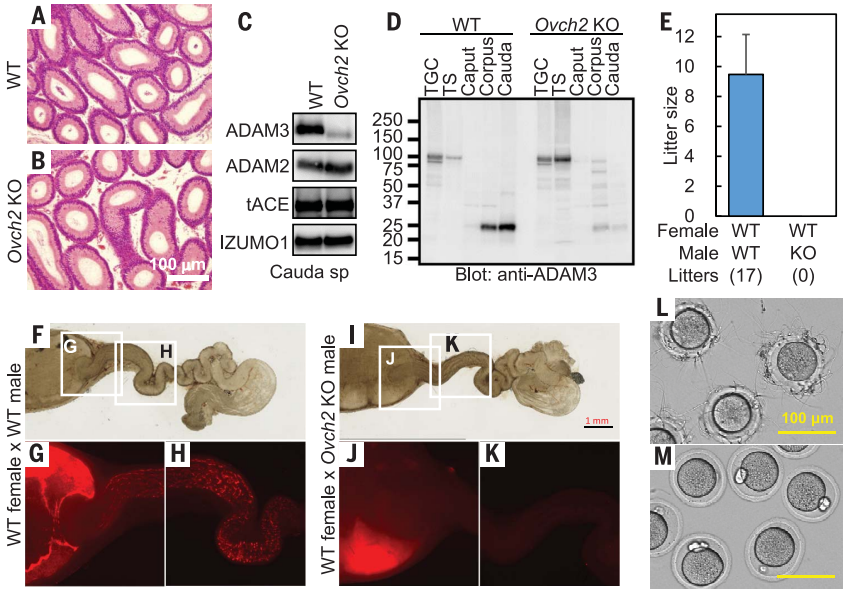


Fig. 3. Epididymal secreted protease OVCH2 is required for ADAM3 processing and male fertility. (**A** and **B**) H&E staining of IS sections of epididymis from (A) WT and (B) *Ovch2* KO mice. Scale bar, 100 μm. (**C**) Immunoblot detection of cauda epididymal sperm lysates. (**D**) Immunoblot detection of ADAM3 in TGC, TS, caput, corpus, and cauda epididymal spermatozoa. (**E**) Litter size analysis of *Ovch2* KO mice. Average and SD are shown. (**F** to **K**) Migration of [(F) to (H)] dsRed2-tagged WT and [(I) to (K)] *Ovch2* KO spermatozoa ejaculated into WT female reproductive tract. (**L** and **M**) Sperm-ZP binding assay with (L) WT and (M) *Ovch2* KO spermatozoa. Scale bars, 100 μm.

transmembrane protein, in cauda epididymal mature spermatozoa (*19*). ADAM3 is expressed in TGCs as a precursor N-linked glycosylated protein of ~100 kD and becomes processed by an unknown mechanism to a mature form of ~25 kD during sperm transit through the epididymis (*20*). We found that precursor ADAM3 is not processed correctly to mature ADAM3 in *Nell2* KO cauda epididymal spermatozoa (Fig. 2, I and J). Expression of other fertility-related sperm proteins was not critically affected in *Nell2* KO testis and spermatozoa (Fig. 2I and fig. S3). Transgenic rescue of *Nell2* KO males with testis-specific NELL2 expression restored sperm ADAM3 processing (fig. S4). These results indicate that *Nell2* KO epididymis is missing a key protease involved in proper ADAM3 processing.

We hypothesized that a luminal environment suitable for the maturation of spermatozoa, including ADAM3 processing, is provided by fully differentiated IS but is impaired in *Nell2* KO epididymis. We examined the expression of genes encoding proteases in caput epididymis by means of RNA-seq analysis (Fig. 2K and fig. S7). With RNA-seq and immunoblot analyses, we found that two proteases, OVCH2 (ovochymase 2) and ADAM28, were strongly expressed in WT caput epididymis (fig. S7) but highly down-regulated in *Nell2* KO caput epididymis at the transcript and protein levels (Fig. 2, K and L). As previously reported with ADAM28 (*21*), recombinant OVCH2 showed protease activity in vitro (fig. S7).

To determine whether epididymal secreted proteases are necessary for ADAM3 processing and male fertility, we generated *Ovch2* KO and *Adam28* KO mice using CRISPR-Cas9 (Fig. 3 and figs. S8 and S9). KO mice lacking the epididymis-specific protease OVCH2 partially reproduced the *Nell2* KO phenotype; although IS differentiation was not compromised in *Ovch2* KO mice (Fig. 3, A and B), ADAM3 processing in the *Ovch2* KO spermatozoa was abnormal (Fig. 3, C and D), and *Ovch2* KO male mice were sterile (Fig. 3E) because of

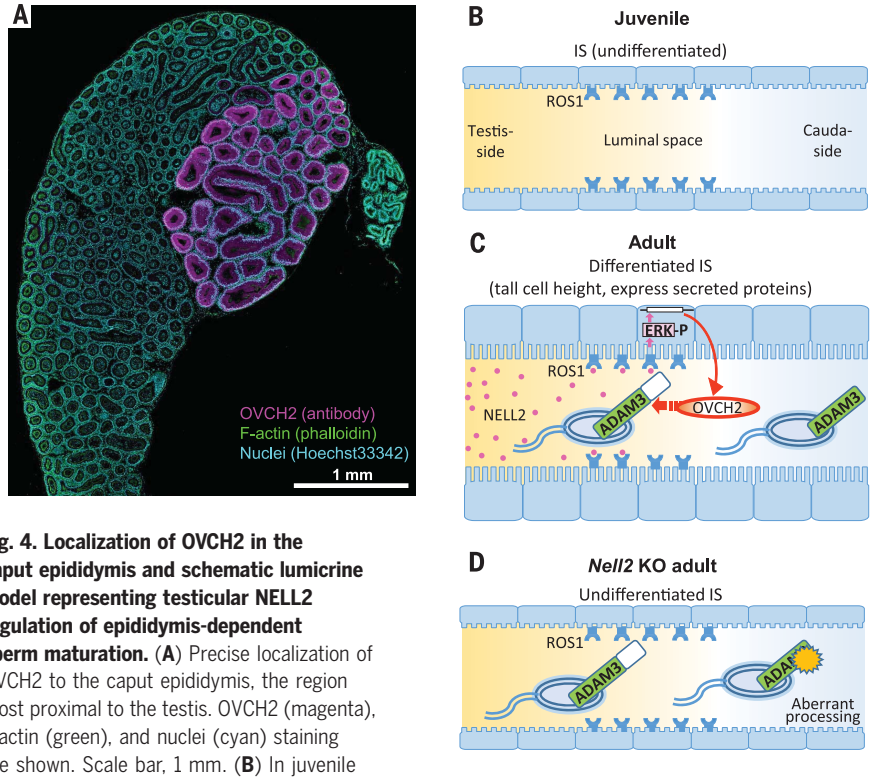


Fig. 4. Localization of OVCH2 in the caput epididymis and schematic lumicrine model representing testicular NELL2 regulation of epididymis-dependent sperm maturation. (**A**) Precise localization of OVCH2 to the caput epididymis, the region most proximal to the testis. OVCH2 (magenta), F-actin (green), and nuclei (cyan) staining are shown. Scale bar, 1 mm. (**B**) In juvenile males, the epididymal IS is undifferentiated. (**C**) During sexual maturation and in adult males, NELL2 is transported from the testis to the epididymis through the luminal space. IS epithelial cells are fully differentiated by the NELL2-ROS1-ERK signaling pathway and induce the secretion of proteases, including OVCH2. OVCH2 subsequently promotes sperm surface ADAM3 processing, which is indispensable for sperm maturation. (**D**) In *Nell2* KO males, IS differentiation and OVCH2 protease secretion are impaired. As a consequence, ADAM3 processing is aberrant, and males are infertile.

inability of the sperm to transit the UTJ (Fig. 3, F to K) and bind to the ZP (Fig. 3, L and M). By contrast, *Adam28* KO males were found to be fertile (fig. S9). Consistent with our model that OVCH2 is downstream of NELL2, *Ovch2* was not only epididymis-specific (fig. S7), OVCH2 protein was exquisitely localized to the IS of the caput epididymis (Fig. 4A), most proximal to the testis. These results indicate that epididymal-specific and secreted protease OVCH2 is required for sperm ADAM3 processing and consequential sperm fertilizing ability.

This study indicates the presence of a lumicrine signal pathway in which a testis-secreted protein (NELL2) signals through a ROS1-pathway to regulate the epididymal IS maturation and subsequent secretion of an epididymis protease (OVCH2) that processes ADAM3 for sperm fertilizing ability (Fig. 4, B to D). NELL2 is produced by developing germ cells in testis and transits through the luminal space to the IS of the epididymis to trigger its differentiation. The differentiated IS then secretes many proteins, including proteases (such as OVCH2) that act on sperm ADAM3 to make spermatozoa fully functional. Because various genes are down-regulated in *Nell2* KO caput epididymis (fig. S7), it is very likely that epididymal luminal factors other than proteases also regulate sperm maturation in a different way but downstream of the NELL2-ROS1 lumicrine system. Likewise, because OVCH2 is required for male fertility, OVCH2 may process other sperm and/or epididymis proteins required for sperm maturation. The lumicrine concept that we have demonstrated here in the male reproductive tract may be applied more broadly to other organs in which luminal flow occurs.

REFERENCES AND NOTES

1. B. T. Hinton, Z. J. Lan, D. B. Rudolph, J. C. Labus, R. J. Lye, *J. Reprod. Fertil. Suppl.* **53**, 47–57 (1998).
2. E. Sonnenberg-Riethmacher, B. Walter, D. Riethmacher, S. Gödecke, C. Birchmeier, *Genes Dev.* **10**, 1184–1193 (1996).
3. B. Xu, A. M. Washington, B. T. Hinton, *Biol. Reprod.* **95**, 15 (2016).
4. A. Naba *et al.*, *Mol. Cell. Proteomics* **11**, 014647 (2012).
5. H. Miyata *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 7704–7710 (2016).
6. F. A. Abdulhaleem *et al.*, *Dev. Neurobiol.* **75**, 494–504 (2015).
7. Y. Muro, M. G. Buffone, M. Okabe, G. L. Gerton, *Biol. Reprod.* **86**, 1–6 (2012).
8. L. Svensson *et al.*, *Mol. Cell. Biol.* **22**, 4366–4371 (2002).
9. C. B. Tempfer, R. M. Moreno, A. R. Gregg, *Biol. Reprod.* **62**, 457–462 (2000).
10. O. Burnicka-Turek *et al.*, *Endocrinology* **150**, 4348–4357 (2009).
11. M. Yamashita *et al.*, *Genes Cells* **13**, 1001–1013 (2008).
12. B. P. Hermann *et al.*, *Cell Rep.* **25**, 1650–1667.e8 (2018).
13. D. Watanabe *et al.*, *FEBS Lett.* **368**, 509–512 (1995).
14. B. Xu, L. Yang, R. J. Lye, B. T. Hinton, *Biol. Reprod.* **83**, 807–817 (2010).
15. H. J. Jun *et al.*, *Endocrinology* **155**, 3661–3673 (2014).
16. T. Nakanishi *et al.*, *FEBS Lett.* **449**, 277–283 (1999).
17. H. Hasuwa *et al.*, *Exp. Anim.* **59**, 105–107 (2010).
18. C. H. Yeung, A. Wagenfeld, E. Nieschlag, T. G. Cooper, *Biol. Reprod.* **63**, 612–618 (2000).
19. Y. Fujihara, H. Miyata, M. Ikawa, *Exp. Anim.* **67**, 91–104 (2018).
20. E. Kim *et al.*, *J. Reprod. Dev.* **50**, 571–578 (2004).

ACKNOWLEDGMENTS

We thank the Biotechnology Research and Development (nonprofit organization) and Department of Experimental Genome Research for the generation of the mutant mice. We thank T. Enomoto for sharing resources and advice. **Funding:** This work was supported by KAKENHI grants JP15K06999 (to D.K.); JP18K14612 (to T.N.); JP16KK0180 (to Y.F.); and JP25112007, JP17H01394, JP19H05750 (to M.I.); AMED JP19 g5010001 and the Takeda Science Foundation (to M.I.); the Eunice Kennedy Shriver National Institute of Child Health and Human Development (R01HD088412 and P01HD087157 to M.M.M. and M.I.), and The Bill and Melinda Gates Foundation (INV-001902 to M.M.M. and M.I.).
Author contributions: D.K., M.Ok., M.M.M., and M.I. designed experiments. D.K., R.Y., T.T., T.N., T.M., K.S., M.K., T.K., Y.F., M.Oz., Z.Y., G.M., and K.M.B. performed experiments. M.H. contributed resources. D.K., M.Ok., M.M.M., and M.I. wrote the manuscript.
Competing interests: The authors declare no competing interests.
Data and materials availability: RNA-seq data obtained from WT caput, corpus, and cauda epididymis have been deposited in the Gene Expression Omnibus (GEO) under accession code GSE138517 (www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE138517). RNA-seq data obtained from WT and *Nell2* KO caput epididymis have been deposited in the GEO under accession code GSE133920 (www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE133920). All other data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6495/1132/suppl/DC1
Materials and Methods
Figs. S1 and S9
Tables S1 and S2
References (*21*–*31*)

View/request a protocol for this paper from *Bio-protocol*.

8 July 2019; accepted 12 April 2020
10.1126/science.aay5134

MOLECULAR BIOLOGY

Barcoded microbial system for high-resolution object provenance

Jason Qian[1,2,3]*, Zhi-xiang Lu[1,2]*, Christopher P. Mancuso[4]*, Han-Ying Jhuang[1]*, Rocío del Carmen Barajas-Ornelas[5]*, Sarah A. Boswell[1,2]*, Fernando H. Ramírez-Guadiana[5], Victoria Jones[1,6]†, Akhila Sonti[4], Kole Sedlack[4]‡, Lior Artzi[5], Giyoung Jung[7], Mohammad Arammash[1], Mary E. Pettit[1], Michael Melfi[1], Lorena Lyon[1], Siân V. Owen[6], Michael Baym[2,6], Ahmad S. Khalil[4,8], Pamela A. Silver[1,8], David Z. Rudner[5], Michael Springer[1,2]§

Determining where an object has been is a fundamental challenge for human health, commerce, and food safety. Location-specific microbes in principle offer a cheap and sensitive way to determine object provenance. We created a synthetic, scalable microbial spore system that identifies object provenance in under 1 hour at meter-scale resolution and near single-spore sensitivity and can be safely introduced into and recovered from the environment. This system solves the key challenges in object provenance: persistence in the environment, scalability, rapid and facile decoding, and biocontainment. Our system is compatible with SHERLOCK, a Cas13a RNA-guided nucleic acid detection assay, facilitating its implementation in a wide range of applications.

Globalization of supply chains has substantially complicated the process of determining the origins of agricultural products and manufactured goods. Determining the origins of these objects can be critical, for example, in cases of foodborne illness, but current labeling technologies are prohibitively labor intensive and easy to subvert (*1*). Tools that label persons or objects passing through a location of interest could also be useful to law enforcement as a complement to fingerprinting and video surveillance (*2*). Microbial communities offer a potential alternative to standard labeling approaches. Any object gradually adopts the naturally occurring microbes present in its environment (*3*, *4*), so it has been suggested that the microbial composition of an object could be used to determine its provenance (*5*). Challenges with this approach include variability of resident microbial community abundance over time; similarities of microbial composition between different locations; and the requirement for extensive, expensive, and time-consuming mapping of natural environments.

To circumvent these challenges, we propose the deliberate introduction and use of synthetic, nonviable microbial spores harboring barcodes that uniquely identify locations of interest (e.g., food production areas). These synthetic spores would offer a sensitive, inexpensive, and safe way to map object provenance provided that several important criteria are met, including: (i) the microbes must be compatible with growth at industrial scale; (ii) the synthetic spores must be biocontained and not viable in the wild to prevent adverse ecological effects; (iii) the synthetic spores must persist in the environment and reliably label objects that pass through it; and (iv) the encoding and decoding of information about object provenance must be rapid, sensitive, and specific. Similar barcoding approaches have been explored previously to model pathogen transmission (*6*, *7*) but did not explicitly address those challenges. Here, we report the barcoded microbial spores (BMS) system, a scalable, safe, and sensitive system that uses DNA-BMS mixtures to permit the determination of object provenance (Fig. 1A).

The BMS system leverages the natural ability of spores to persist for long periods in the

[1]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA. [2]Laboratory of Systems Pharmacology, Harvard Medical School, Boston, MA 02115, USA. [3]Biological and Biomedical Sciences Program, Harvard Medical School, Boston, MA 02115, USA. [4]Department of Biomedical Engineering and Biological Design Center, Boston University, Boston, MA 02215, USA. [5]Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA 02115, USA. [6]Department of Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA. [7]Synthetic Biology Center, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [8]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115, USA.
*These authors contributed equally to this work.
†Present address: Quanterix, Billerica, MA 01821, USA.
‡Present address: The University of Illinois College of Medicine, Rockford, IL 61107, USA.
§Corresponding author. Email: michael_springer@hms.harvard.edu

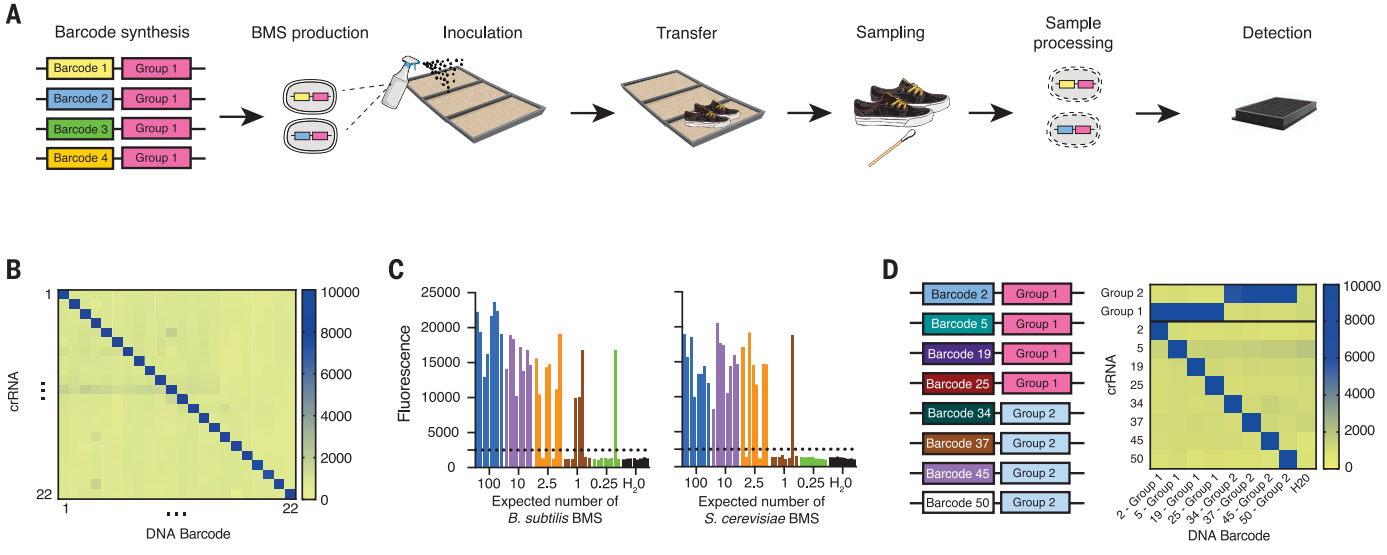


Fig. 1. BMS can be specifically and sensitively detected. (**A**) Schematic of the BMS application and detection pipeline. (**B**) Heatmap of endpoint fluorescence values from in vitro SHERLOCK reactions of all combinations of 22 barcodes and 22 crRNAs assessing specificity of each barcode-crRNA pair. (**C**) Detection limit of *B. subtilis* and *S. cerevisiae* BMS by SHERLOCK (each of the eight biological replicates for each spore concentration are shown). Spore numbers are calculated on a per-reaction basis. (**D**) Heatmap of endpoint fluorescence values from in vitro SHERLOCK reactions testing the specificity of four barcodes for group 1 crRNA and four barcodes for group 2 crRNA as detected by either specific or group crRNA.

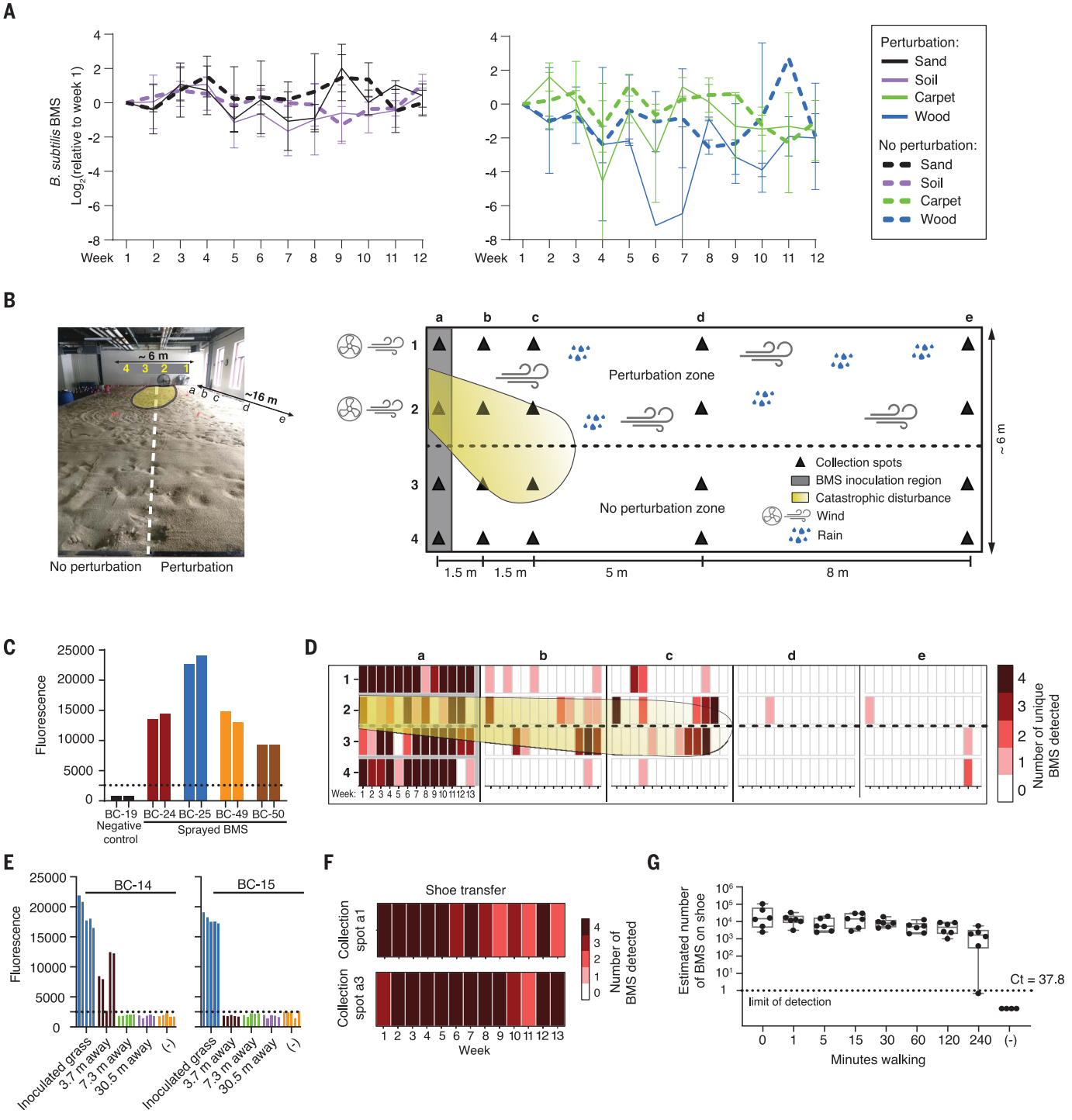
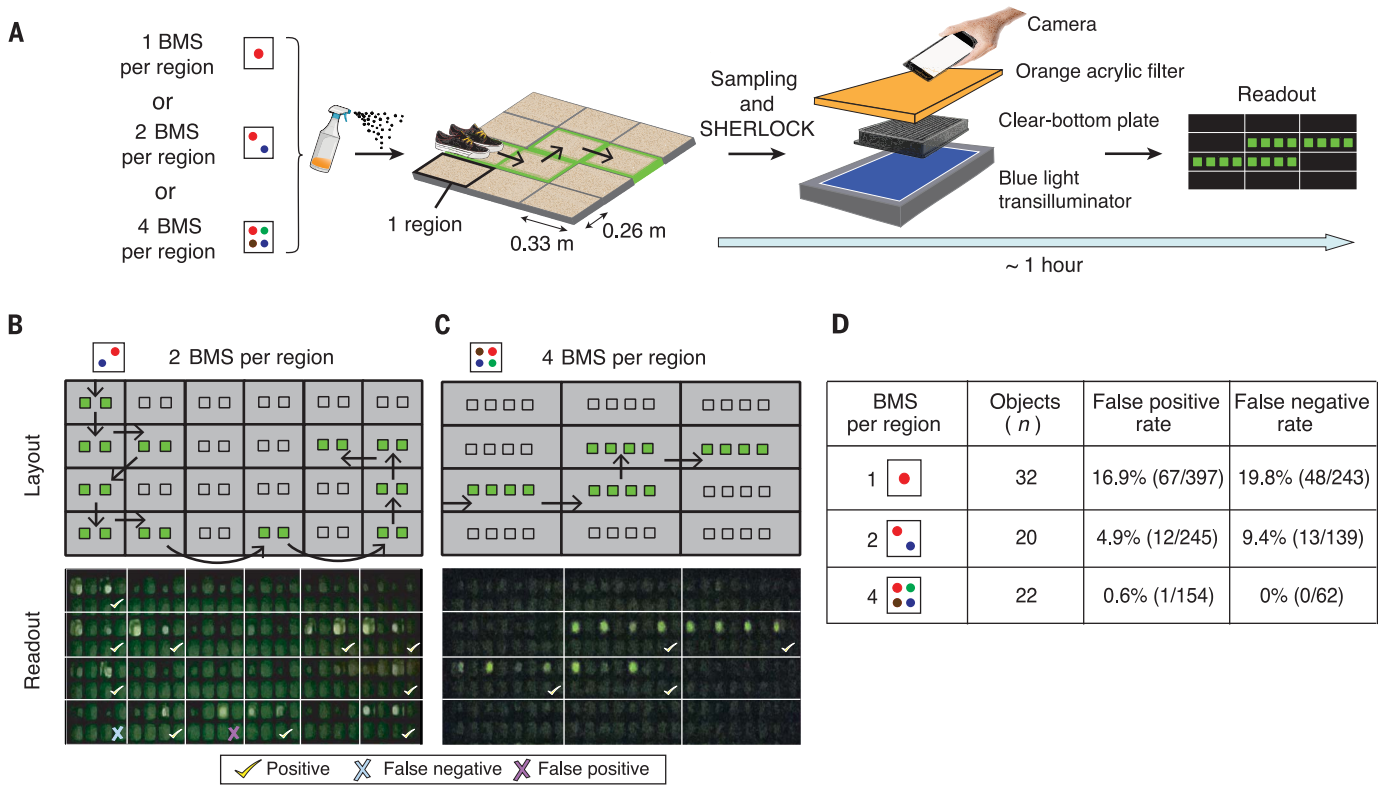


Fig. 2. Persistence, transferability, and maintenance of BMS. (**A**) BMS persisted on sand, soil, carpet, and wood over 3 months on ~1-m^2 test surfaces in incubators. The *y*-axis shows the *B. subtilis* BMS number relative to week 1 levels. Bars represent standard deviations. Perturbations were simulated wind, rain, vacuuming, or sweeping. (**B**) Photograph and schematic of a large-scale (~100-m^2) sandpit. *B. subtilis* BC-24 and BC-25 BMS and *S. cerevisiae* BC-49 and BC-50 BMS were inoculated in the shaded gray region in the diagram. The yellow shadow represents the area over which the top 5 cm of sand from a 1.5-m^2 area of the inoculation region was redistributed after a large fan fell over. (**C**) SHERLOCK signal from the four BMS from the inoculated region "a" in the large-scale experiment. Dashed line is the threshold for positive calls. BC-19 is the negative control. (**D**) BMS persist at collection point "a" (within the inoculated region) and do not spread to collection points "d" or "e." Heatmap depicts the number of BMS (out of four) detected by SHERLOCK at each collection point over 13 weeks. (**E**) BMS persist on grass in an outdoor environment for at least 5 months. The grass region was inoculated with *B. subtilis* BC-14 and 15 BMS. Samples from actual BMS-inoculated region 3.7, 7.3, and 30.5 m away from the inoculated region were tested by SHERLOCK using crRNA 14 and 15 (each of the five biological replicates for each grass region are shown). (**F**) BMS were transferred onto shoes by walking in the inoculated region "a" in the sandpit and were detected by SHERLOCK. (**G**) Abundance of BC-25 BMS on shoes after up to 240 min of walking on noninoculated outdoor areas. The *y*-axis shows the BMS count based on the qPCR standard curve.



D

BMS per region	Objects (*n*)	False positive rate	False negative rate
1	32	16.9% (67/397)	19.8% (48/243)
2	20	4.9% (12/245)	9.4% (13/139)
4	22	0.6% (1/154)	0% (0/62)

Fig. 3. Determining object provenance using BMS and field-deployable detection. (**A**) Schematic of experiment design and field-deployable method to determine previous locations of an object. Each region was inoculated with one, two, or four nonredundant BMS. Green outlines indicate the object path through a subset of the regions. SHERLOCK reactions were imaged using a mobile phone camera to photograph the reaction plate through a filter and under portable blue light illumination. (**B** and **C**) (Top) Path of an object overlaid on a reaction plate. (Bottom) Photograph of SHERLOCK reaction plate overlaid with correct or incorrect calls. The call for each region is denoted by color (yellow check: true positive, blue cross: false negative, purple cross: false positive). (**D**) Statistics for SHERLOCK provenance predictions of objects traversing regions inoculated with one, two, or four BMS per region. The false-positive and false-negative rates for one and two BMS per region were based on the criteria of one or more positive calls; the rates for four specific BMS per region were based on the criteria of two or more positive calls.

environment without growth (*8*). We designed nonredundant DNA barcodes and integrated them into the genomes of *Bacillus subtilis* and *Saccharomyces cerevisiae* spores, creating a set of BMS that can be used combinatorially to provide a nearly infinite set of identification codes. The BMS can be manufactured at scale using standard cloning and culturing techniques, inoculated to surfaces by spraying, and transferred to objects that come into contact with the inoculated surface. To identify barcodes, BMS sampled from objects can be lysed and decoded with a range of methods including SHERLOCK, a recombinase polymerase amplification (RPA) method, coupled with a Cas13a-based nucleic acid detection assay (*9*), quantitative polymerase chain reaction (qPCR), and sequencing (Fig. 1A and fig. S1A).

The BMS are designed not to affect the native environment into which they are applied. First, we used auxotrophic strains that require amino acid supplementation for growth. Second, we made the cells germination deficient. For *B. subtilis* spores, we deleted the genes encoding the germinant receptors and the genes that encode the cell wall lytic enzymes required to degrade the specialized spore cell wall. Incubation of $>10^{12}$ spores generated from this mutant strain showed they were unable to form colonies or grow in rich medium and remained stable and nongerminating at room temperature for >3 months (fig. S2, A and C). For *S. cerevisiae*, we boiled spores for 30 min to heat-kill vegetative cells and spores before application. Incubation of $>10^{8}$ boiled spores on rich medium yielded no colonies (fig. S2, B and D to F). All antibiotic resistance cassettes used to generate the BMS were removed by site-specific recombination to prevent horizontal gene transfer of resistance genes to other organisms in the environment. Finally, the inserted barcode does not encode any gene and should not confer any fitness advantage if horizontally transferred. We tracked the microbiome of soil samples and found that inoculation with BMS had insignificant effects on the microbiome compared with natural changes over time or in response to watering (fig. S3). We also note that *B. subtilis* and *S. cerevisiae* are both commonly found in environmental and food samples.

Multiple BMS can be applied and then decoded simultaneously. We designed a series of tandem DNA barcodes, each with a Hamming distance of >5, allowing $>10^{9}$ possible individual barcodes. To test the specificity of our barcode design in a field-deployable system, we constructed 22 barcodes and their matching CRISPR RNAs (crRNAs) and assayed all permutations in vitro using SHERLOCK. All 22 crRNAs clearly distinguished the correct barcode target (Fig. 1B). To scale the system, we devised a rapid and facile method to screen a large number of barcodes and crRNAs in parallel to eliminate those with cross-reactivity or background. We validated and performed pooled *n-1 barcode* RPA reactions in vitro with corresponding crRNA and water RPA controls testing 94 crRNA-barcode pairs, eliminating 17 for high background and seven

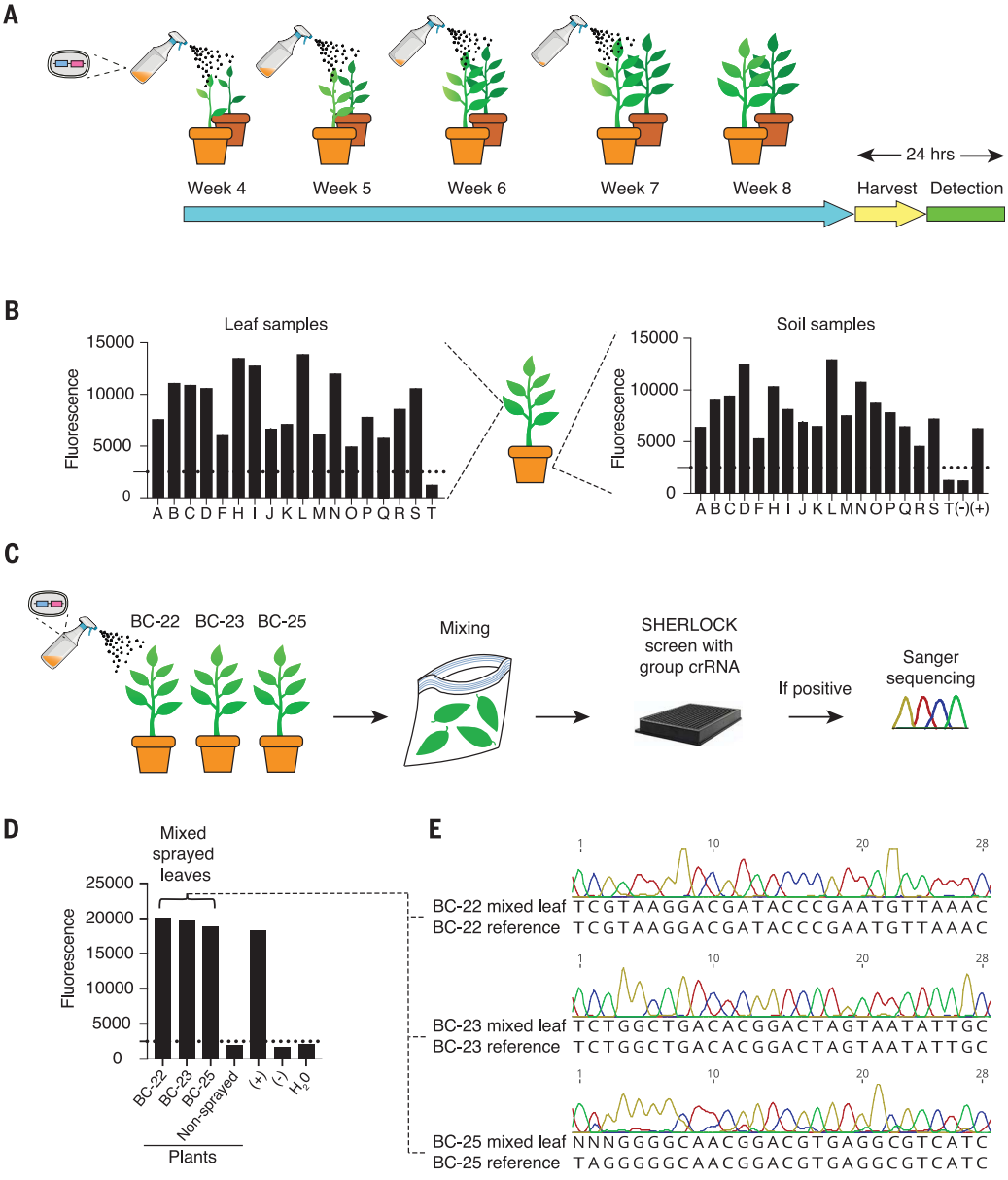


Fig. 4. Determining the provenance of produce using BMS. (**A**) Eighteen plants were inoculated with distinct *B. subtilis* BMS and inoculated once per week (four times total). (**B**) Detection of BMS on plants and soil after harvesting by SHERLOCK with a group crRNA. The *y*-axis shows endpoint fluorescence values. Plants A to S were sprayed with BMS; plant T was not sprayed. (−), negative control without DNA template; (+), positive control from DNA. Dashed line is the threshold for positive calls. (**C**) Leaves from plants inoculated with different BMS were mixed together, SHERLOCK was used to confirm the presence of the BMS, and then Sanger sequencing was used to identify the origin of each leaf. (**D**) Leaves were screened for the presence of BMS by SHERLOCK using a group crRNA. The *y*-axis shows endpoint fluorescence values. (+), group 2 positive DNA; (−), group 1 DNA; (H_2O), water control. (**E**) Sanger sequencing identified the plant of origin of the mixed leaves.

for cross-reactivity (fig. S1, B and C). To test sensitivity and specificity in vivo, we integrated 57 barcodes into *B. subtilis* and 11 into *S. cerevisiae*. We developed an efficient spore lysis protocol using heat and sodium hydroxide (fig. S4), which allowed us to achieve near single-spore resolution for detection using SHERLOCK (Fig. 1C). In vivo and in vitro specificity screenings of crRNA-barcode pairs gave similar results (fig. S1, C and D). In addition, our barcodes are tandemly designed with a specific sequence and a shared group sequence (fig. S1A) to aid in high-throughput detection settings where only a subset of samples contains the BMS of interest. The group sequence is compatible with field-deployable detections and can be used to determine whether a BMS of interest is present before using a second assay to specifically identify the BMS (Fig. 1D). This two-step process solves the throughput limitations of field-deployable detection and lowers the costs of sequencing.

The BMS system is robust and can function on different surfaces in simulated real-world environments. First, in ~1-m^2-scale experiments (fig. S5A and table S6), we used qPCR to detect and quantify BMS directly from surface samples or surface swabs. We found that BMS persisted on sand, soil, carpet, and wood surfaces for at least 3 months with little to no loss over time (Fig. 2A and fig. S5, B and C). Notably, multiple tested perturbations (e.g., simulated wind, rain, vacuuming, or sweeping; see fig. S5A) did not reduce our ability to detect BMS from the surface. Second, we constructed an ~100-m^2 indoor sandpit (Fig. 2B and fig. S6), inoculated one region with BMS (Fig. 2C), and were able to readily detect the BMS for 3 months using SHERLOCK (Fig. 2D and fig. S7). Perturbations did not cause appreciable spreading to noninoculated areas (Fig. 2D and fig. S7); even a catastrophic disturbance in which a fan fell into the sand, only spread the BMS several meters (fig. S6). In an outdoor environment, BMS inoculated on grass was still detectable after 5 months of exposure to natural weather, with minimal spreading outside of the inoculated region (Fig. 2E). This is consistent with low levels of re-aerosolization reported for other spore-forming Bacilli (*10*).

The BMS can be transferred onto objects that pass through test environments. In ~1-m^2-scale testing, BMS could be transferred onto

rubber or wooden objects simply by placing them on the BMS-inoculated surface for several seconds, yielding up to ~100 spores/μl of reaction input (fig. S5D). At ~100-m^2 scale, BMS were reliably transferred onto shoes worn in the inoculated sandpit (Fig. 2F and fig. S8). Furthermore, the BMS transferred onto shoes could still be detected even after walking on noninoculated surfaces for several hours, although BMS counts decreased by twofold with 2 hours of walking as quantified by qPCR (Fig. 2G and fig. S9). We were unable to detect spores on noninoculated surfaces after walking on them with shoes that had traveled through BMS-inoculated regions (fig. S10). We conclude that the BMS can persist in the environment without significant spreading, are transferable onto objects that pass through the environment, are retained on these objects, and can be sensitively and specifically detected using SHERLOCK.

The BMS system can be used to label specific locations of interest to determine whether a person or object has passed through them. We divided different surfaces into grids; inoculated each grid region with one, two, or four nonredundant BMS (Fig. 3A); and traversed them with different test objects (e.g., shoes). To mimic in-field deployment, we used a portable light source, an acrylic filter, and a mobile phone camera to image the SHERLOCK readout (Fig. 3A and fig. S11A) and to determine object provenance (Fig. 3, B and C, and figs. S11 to S13). Provenance could be determined in the field within ~1 hour from sample collection. To evaluate the sensitivity and specificity of our system for determining an object's provenance, we considered different criteria for classification with varying numbers of BMS per region. Object provenance could be determined with a 0.6% (1/154) false-positive rate and a 0% (0/62) false-negative rate if regions were inoculated with four BMS based on the criteria of two or more positive BMS calls. Inoculating with only one or two BMS per region still permitted object provenance to be determined, albeit at the higher error rates (Fig. 3D and fig. S14). Provenance could be determined on all four surface types tested (sand, soil, carpet, and wood) (fig. S13); further validation will be needed to determine error rates in real-world environments. This experiment demonstrates that the BMS can be used to determine object provenance at meter-scale resolution, which would be extremely difficult to achieve using natural microbiome signatures (*11*).

The BMS system offers a flexible and comprehensive approach to determining food provenance. Foodborne illness is a global health issue with an estimated 48 million cases each year in the United States alone (*12*). There is an urgent need for rapid methods for identifying the source of food contamination; current approaches often take weeks and are costly because of the complex modern market chain (*13*). Plants inoculated with *B. subtilis* BMS allowed us to map laboratory-grown leafy plants back to the specific pot in which they were grown (Fig. 4, A and B, and fig. S15). BMS were inoculated four times, beginning 1 week after the first set of leaves appeared, to match recommended inoculation protocols for *Bacillus thuringiensis* (*Bt*) spores, a U.S. Food and Drug Administration–approved biocide that is widely used in agriculture (*14*). One week after the final BMS inoculation, a leaf and a soil sample from each pot were harvested and tested using SHERLOCK. All of the samples were positively detected except for the two plants that had received variant group barcode sequences, demonstrating the specificity of detection (Fig. 4B and fig. S15, B and C). Using Sanger sequencing, we then identified the pot in which each plant was grown for all 18 BMS (fig. S15D). The process from DNA extraction to sequence identification took <24 hours. This time frame could likely be shortened to hours with massively parallelized hybridization-based detection (*15*).

Cross-association of BMS-inoculated plants does not compromise the determination of provenance. To simulate cross-association that could occur during food processing, we mixed leaves from plants that were inoculated with a specific BMS per plant. Unlike the other surfaces that we inoculated, BMS-inoculated plants did not transfer as easily to objects that came into contact with the plants (fig. S5 versus fig. S16, A and B). Although there was detectable transfer between leaves, the amount of transfer still allowed Sanger sequencing to cleanly determine the origin of each leaf (Fig. 4, C to E, and fig. S16, C and D).

Bt could be used to determine food provenance. We used *Bt* spores applied during farming as a surrogate to test whether BMS would persist through conditions of a real-world food supply chain. For plants of known *Bt* inoculation status, we correctly identified all *Bt*-positive and -negative plants (38 total plants) (fig. S17). Furthermore, we detected *Bt* on 10 of 24 store-bought produce items of a priori unknown *Bt* status (figs. S17B and S18). Unexpectedly, BMS and *Bt* spores remained detectable even after washing, boiling, frying, and microwaving (fig. S19), highlighting the potential to determine provenance from cooked foods. These results show the potential for using the BMS system to determine produce provenance.

Our work shows how rationally engineered microbial spores can be manufactured in a high-throughput manner to provide a new solution to the object provenance problem. We have shown that BMS (i) persist in the environment; (ii) do not spread out of the inoculation area; (iii) transfer from soil, sand, wood, and carpet to contacting objects; and (iv) permit sensitive and rapid readout using laboratory and field-deployable methods. The ability to rapidly label objects and determine their provenance in real-world environments has a broad range of applications across agriculture, commerce, and forensics (*2*). Preliminary data suggest that BMS could work across various environments, although extensive validation in a wider range of real-world conditions is needed. Future iterations of our BMS system could be engineered for limited propagation and actively contained for use in highly trafficked areas. This system could also provide time-resolved information about location history, making it useful for an even wider range of applications.

REFERENCES AND NOTES

1. P. M. (Nel) Wognum, H. Bremmers, J. H. Trienekens, J. G. A. J. van der Vorst, J. M. Bloemhof, *Adv. Eng. Inform.* **25**, 65–76 (2011).
2. J. Gooch, B. Daniel, V. Abbate, N. Frascione, *Trends Analyt. Chem.* **83**, 49–54 (2016).
3. S. Lax *et al*., *Science* **345**, 1048–1052 (2014).
4. C. Jiang *et al*., *Cell* **175**, 277–291.e31 (2018).
5. S. Lax *et al*., *Microbiome* **3**, 21 (2015).
6. P. Buckley *et al*., *Appl. Environ. Microbiol.* **78**, 8272–8280 (2012).
7. P. A. Emanuel *et al*., *Appl. Environ. Microbiol.* **78**, 8281–8288 (2012).
8. N. Ulrich *et al*., *PLOS ONE* **13**, e0208425 (2018).
9. J. S. Gootenberg *et al*., *Science* **356**, 438–442 (2017).
10. A. H. Bishop, C. M. O'Sullivan, A. Lane, M. C. Butler Ellis, W. J. Sellors, *Lett. Appl. Microbiol.* **64**, 364–369 (2017).
11. R. I. Adams, A. C. Bateman, H. M. Bik, J. F. Meadow, *Microbiome* **3**, 49 (2015).
12. E. Scallan, P. M. Griffin, F. J. Angulo, R. V. Tauxe, R. M. Hoekstra, *Emerg. Infect. Dis.* **17**, 16–22 (2011).
13. C. W. Hedberg *et al*., *Emerg. Infect. Dis.* **14**, 311–313 (2008).
14. G. Sanahuja, R. Banakar, R. M. Twyman, T. Capell, P. Christou, *Plant Biotechnol. J.* **9**, 283–300 (2011).
15. M. Sarwat, M. M. Yamdagni, *Crit. Rev. Biotechnol.* **36**, 191–203 (2016).
16. B.-M. Koo *et al*., *Cell Syst.* **4**, 291–305.e7 (2017).
17. A. J. Meeske *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 6437–6442 (2015).
18. P. Schaeffer, J. Millet, J. P. Aubert, *Proc. Natl. Acad. Sci. U.S.A.* **54**, 704–711 (1965).
19. H. C. Lim *et al*., *PLOS Genet.* **15**, e1008284 (2019).
20. A. A. Horwitz *et al*., *Cell Syst.* **1**, 88–96 (2015).
21. D. M. Gohl *et al*., *Nat. Biotechnol.* **34**, 942–949 (2016).
22. E. Bolyen *et al*., *Nat. Biotechnol.* **37**, 852–857 (2019).
23. B. J. Callahan *et al*., *Nat. Methods* **13**, 581–583 (2016).
24. Q. Chang, Y. Luan, F. Sun, *BMC Bioinformatics* **12**, 118 (2011).
25. L. A. Klobutcher, K. Ragkousi, P. Setlow, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 165–170 (2006).
26. P. S. V. Devi, T. Ravinder, C. Jaidev, *J. Invertebr. Pathol.* **88**, 163–168 (2005).

ACKNOWLEDGMENTS

We thank participating laboratory members for useful feedback, O. Mazor of the HMS Research Instrumentation Core Facility for technical consultations and instrument design and fabrication, and K. Rhea and P. Buckley for BMS-labeled grass samples. **Funding:** This work was supported by DARPA BRICS grant no. HR001117S0029. J.Q. is supported by an NSF GRFP. **Author contributions:** J.Q., Z.L., C.P.M., H.Y.J., R.C.B.-O., and S.A.B. conceived the study and performed most key experiments and data analysis (supervised by M.S.). J.Q., Z.L., C.P.M., H.Y.J., R.C.B.-O., S.A.B., and F.H.R.-G. wrote the paper (with M.S. and D.Z.R.). V.J., A.S., K.S., L.A., G.J., M.A., M.E.P., M.M., L.L., and S.V.O. assisted with technical

experiments. S.A.B. purified Cas13 proteins. R.C.B.-O. and F.H.R.-G. designed and constructed *B. subtilis* strains and purified spores (supervised by D.Z.R.). D.Z.R., P.A.S., A.S.K., and M.B. provided critical insights. All authors reviewed the manuscript. **Competing interests:** U.S. provisional patent application no. US 62/958,512 has been filed on behalf of M.S. and the President and Fellows of Harvard College. **Data and materials availability:** All data are available in the article or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6495/1135/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S19
Tables S1 to S6

View/request a protocol for this paper from *Bio-protocol*.

13 December 2019; accepted 31 March 2020
10.1126/science.aba5584

PLASTIC POLLUTION

Seafloor microplastic hotspots controlled by deep-sea circulation

Ian A. Kane[1]*, Michael A. Clare[2], Elda Miramontes[3,4], Roy Wogelius[1], James J. Rothwell[5], Pierre Garreau[6], Florian Pohl[7]

Although microplastics are known to pervade the global seafloor, the processes that control their dispersal and concentration in the deep sea remain largely unknown. Here, we show that thermohaline-driven currents, which build extensive seafloor sediment accumulations, can control the distribution of microplastics and create hotspots with the highest concentrations reported for any seafloor setting (190 pieces per 50 grams). Previous studies propose that microplastics are transported to the seafloor by vertical settling from surface accumulations; here, we demonstrate that the spatial distribution and ultimate fate of microplastics are strongly controlled by near-bed thermohaline currents (bottom currents). These currents are known to supply oxygen and nutrients to deep-sea benthos, suggesting that deep-sea biodiversity hotspots are also likely to be microplastic hotspots.

Plastic pollution has been observed in nearly all environments on Earth (*1*) and across all of its oceans (*2*–*4*). The effects of plastic pollution on marine ecosystems and implications for human health are of growing concern, as more than 10 million tonnes of plastic enter the global ocean each year (*4*–*6*). Converging surface currents in oceanic gyres are responsible for the global distribution of plastics on the ocean surface (*2*, *3*). These gyres effectively concentrate positively buoyant plastics into the now-infamous "garbage patches" (*2*, *3*). However, sea surface accumulations only account for ~1% of the estimated global marine plastic budget (*3*, *4*, *7*, *8*). Most of the remaining 99% of plastic ends up in the deep sea (*7*–*9*) (Fig. 1A). A considerable proportion [estimated at 13.5% (*8*)] of the marine plastic budget occurs as microplastics: small (<1 mm) fragments and fibers (*10*, *11*) that originate as manufactured particles (*12*, *13*) or are derived from synthetic textiles (*14*) or the breakdown of larger plastic debris (*15*). It has been shown that larger plastic debris may be associated with dense down-canyon flows in the Mediterranean (*16*). The seafloor is a globally important sink for plastics; however, the physical controls on the distribution of microplastics and the effectiveness of their sequestration once deposited at the seafloor remain unclear (*7*, *10*, *17*–*23*). Owing to their small size, microplastics can be ingested by organisms across all trophic levels, enabling transfer of harmful toxic substances (*9*, *10*, *22*). Therefore, determining where microplastics accumulate and their availability for incorporation into the food chain is fundamental to understanding threats to globally important deep-seafloor ecosystems (*24*).

Rather than corresponding to the extent of overlying surface garbage patches, microplastics on the deep seafloor are preferentially focused within distinct physiographic settings (*7*, *19*). Submarine canyons and deep-ocean trenches, which are foci for episodic yet powerful gravity flows, appear to be microplastic hotspots (*7*, *20*, *25*–*27*) (Fig. 2A). This physiographic bias suggests that the transfer of microplastics to and across the deep seafloor is therefore not solely explained by vertical settling from surface gyres. It is likely that the role of deep-sea currents in the dispersal and concentration of microplastics is similar to that of surface currents (*26*, *28*), yet a paucity of contextual data (e.g., bathymetric, oceanographic, and sedimentological) hinders the linkage of physical transport processes to the distribution and ultimate fate of microplastics. Thermohaline currents acting on the seafloor are one of the most important processes for the deep-sea transport of fine-grained particles and build some of the largest sediment accumulations on our planet [called contourite drifts (*29*)] (Fig. 1B), but their role in sequestering microplastics remains unknown.

Here, we link microplastic pollution on the seafloor to bottom currents by integrating high-resolution geophysical data, sediment sampling, microplastics analysis, and numerical modeling. The Tyrrhenian Sea was selected as the study area because (i) the dimensions and grain size of its physiographic elements are broadly comparable to those of many global settings (*29*–*32*); (ii) its ocean circulation patterns and velocities are comparable to currents globally (*31*, *33*); (iii) its plastic input volumes and locations are well constrained (*34*); and (iv) high-resolution seafloor and ocean circulation data afford the spatial and temporal context to investigate our key questions. We analyze data from the Tyrrhenian Sea, where ocean water circulation is driven by the East Corsican Current and its return branch (Figs. 1C and 2E), which reach local velocities of >0.4 m s^{-1} near the surface and >0.2 m s^{-1} near the seafloor (*30*). The strongest bottom currents generally occur between 600 and 900 m water depth, where they actively sculpt extensive muddy contourite drifts (Fig. 1D) [<10 km wide and up to hundreds of meters thick (*31*)]. The continental shelf is indented by the Caprera slope channel system, which extends downslope to the Olbia basin (*32*) (Fig. 1C). Terrestrial sediment is delivered to the shelf by high-gradient rivers passing through rural, urban, and industrial catchments and accounts for ~80% of the marine plastic budget in the region, with the remainder from shipping and fishing activities (*6*–*8*, *34*, *35*, *36*) (Fig. 1, A and C). In this study, we address three questions: How important are bottom currents for the dispersal and accumulation of microplastics on the deep seafloor? How do variations in bottom current intensity control the spatial distribution of microplastics at the seafloor? And how efficiently are microplastics sequestered after their emplacement at the seafloor?

All seafloor samples were found to contain microplastics (Fig. 3B and table S1), as verified with optical microscopy and Fourier transform infrared (FTIR) spectroscopy (fig. S2). Microplastics presented in two forms, as fibers (70 to 100%) and as fragments (0 to 30%) (Fig. 2, B and C). Microplastic concentration in the Tyrrhenian Sea includes the highest values yet recorded from the deep seafloor (Fig. 2A): up to 182 fibers and nine fragments per 50 g of dried sediment (191 total pieces per 50 g in core 6, equivalent to ~1.9 million pieces per square meter) were recorded in the contourite drift at the base of the northeast Sardinian continental slope (Fig. 3, B and C). This concentration exceeds the highest levels previously recorded, including those from deep-sea trenches, and is more than double that documented in submarine canyons (*27*, *37*–*39*) (Fig. 2A). As contourite drifts occur on most of Earth's continental margins (*29*) (Fig. 1B), the high concentrations recorded here strongly suggest that these drifts are globally important repositories for microplastics.

In our study area there is no relationship between microplastic concentrations and distance from terrestrial plastic sources (Fig. 2B). Samples from the continental shelf (38 fibers and 3 fragments; core 9) and upper slope (8 fibers and 1 fragment; core 11) have some of the lowest concentrations reported in the study area. Instead, we show that microplastics are focused within a water depth range of

[1]School of Earth and Environmental Sciences, University of Manchester, Manchester M13 9PL, UK. [2]National Oceanography Centre, University of Southampton Waterfront Campus, Southampton SO14 3ZH, UK. [3]Faculty of Geosciences, University of Bremen, 28359 Bremen, Germany. [4]MARUM-Center for Marine Environmental Sciences, University of Bremen, 28359 Bremen, Germany. [5]Department of Geography, University of Manchester, Manchester M13 9PL, UK. [6]IFREMER, Univ. Brest, CNRS UMR 6523, IRD, Laboratoire d'Océanographie Physique et Spatiale (LOPS), IUEM, 29280, Plouzané, France. [7]Department of Earth Sciences, Durham University, Durham DH1 3LE, UK.
*Corresponding author. Email: ian.kane@manchester.ac.uk

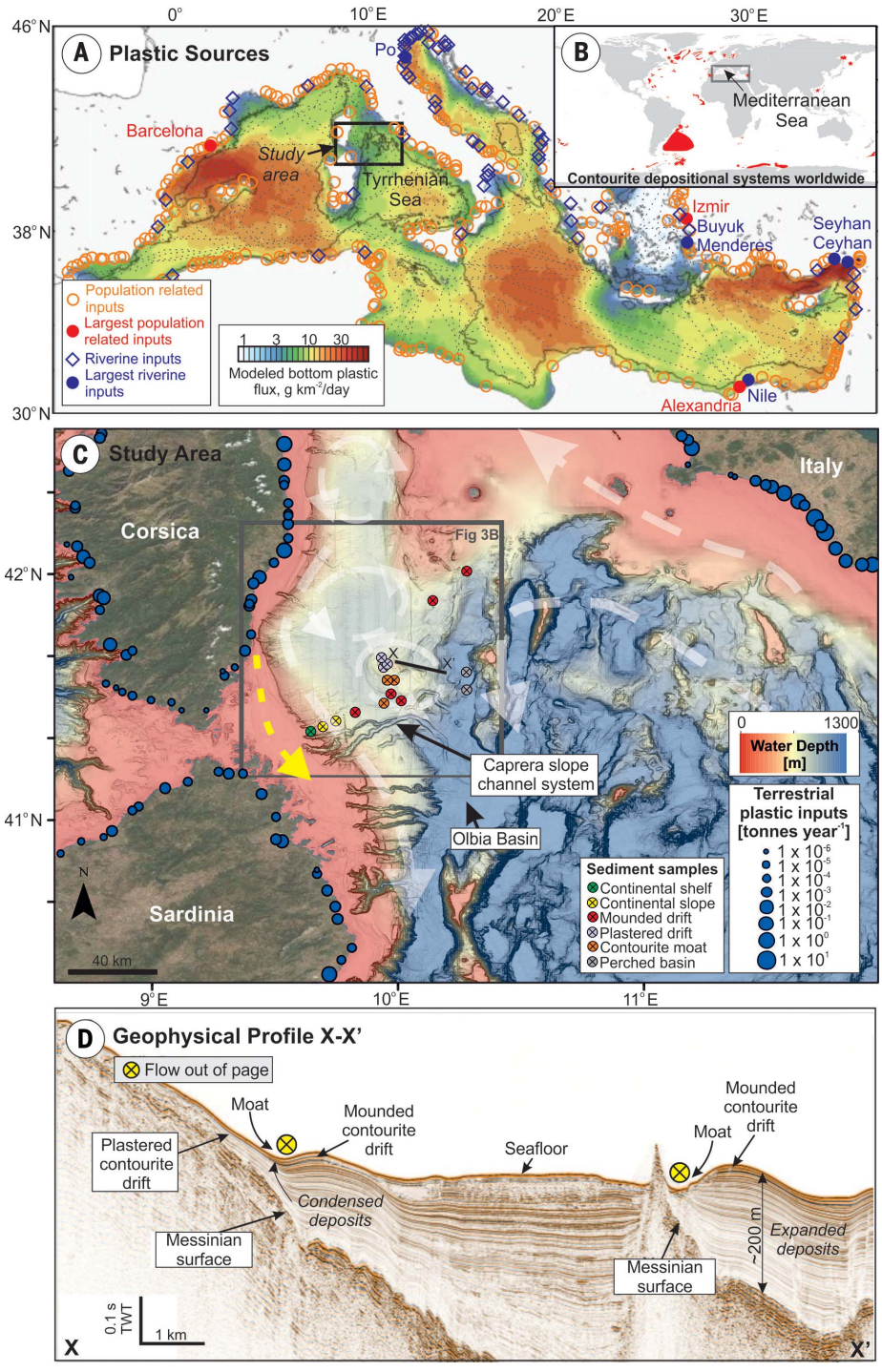


Fig. 1. Plastic sources and ocean circulation affecting the Tyrrhenian Sea. (**A**) Location of study area in the Tyrrhenian Sea, annotated with published terrestrial (~80%) and maritime (fishing and shipping; ~20%) plastic sources (*34*). Terrestrial input sources shown as circles and diamonds. Vessel traffic and shipping lanes shown as dashed lines. Modeled (and hence inferred) plastic debris fluxes on the Mediterranean seafloor illustrated as colored shading. The period modeled was 2013–2017, assuming vertical settling from surface distributions using a two-dimensional Lagrangian model (*34*). Inferred values in the northern Tyrrhenian Sea are <7 g km^{-2} day^{-1}, which are low compared with those of the wider Mediterranean Sea. (**B**) Global distribution of documented contourite depositional systems (*29*) shown in red. (**C**) Seafloor bathymetry of the northern Tyrrhenian Sea annotated with documented terrestrial plastic input sources (*6*), named physiographic features (*32*), and seafloor sediment samples analyzed in this study (colored according to physiographic domain). The regional pattern of thermohaline-driven currents near the seafloor is shown by white arrows. Along-shore drift on the Corsican and Sardinian continental shelves is shown by a yellow dashed arrow. Line X–X′ shows the location of the multichannel seismic line in (**D**), which illustrates the depositional features that have developed as a result of bottom currents, including the formation of thick mounded drifts, and inhibited sediment accumulation in moats. TWT, two-way travel time.

600 to 900 m, where bottom currents form seafloor gyres and have the greatest interaction with the seafloor (Fig. 2C). The influence of these currents and the complex topographic relief (Fig. 2E) result in spatial variations in the shear stress exerted on the seafloor, as determined from hydrodynamic modeling (Fig. 3B). These variations in shear stress explain the localized seafloor distribution of microplastic particles, which typically have lower densities than silt- and sand-forming minerals and therefore are more easily entrained (*39*), accounting for depleted levels of microplastic in certain physiographic domains and concentration in others (Figs. 2D and 3C). The lowest concentrations of microplastics are found in contour-parallel moats, which are foci for erosion and/or nondeposition (e.g., 28 fibers and 3 fragments per 50 g in core 16) (Fig. 3C). Higher concentrations occur on the adjacent mounded drift (e.g., 86 fibers and 1 fragment in core 8) and also in other mounded drift accumulations (e.g., 88 fibers and 6 fragments in core 2) (Fig. 3, B and C).

Although microplastic abundance is generally higher where bottom currents occur, it appears that there is a threshold bed shear stress above which microplastics no longer become concentrated at the seafloor. Modeling of particle transport under the ranges of shear stresses determined from the hydrodynamic model indicates that microplastics are likely to be remobilized and potentially

transported as bedload at shear stresses in excess of 0.03 to 0.04 N m^{-2} (Figs. 3A and 4). Areas of higher shear stress are observed on the shelf break, upper continental slope, and particularly in contourite moats, where the bottom current is most vigorous (Fig. 3B). We propose that such contour currents are effective agents for the transport of microplastics and that, while microplastics are generally flushed along-slope through contourite moats, they preferentially accumulate in the adjacent contourite drifts, forming anomalously concentrated microplastic hotspots. This appears to be particularly true for mounded contourite drifts, which are often sites of extremely high sediment accumulation relative to the rest of a continental slope (*29*).

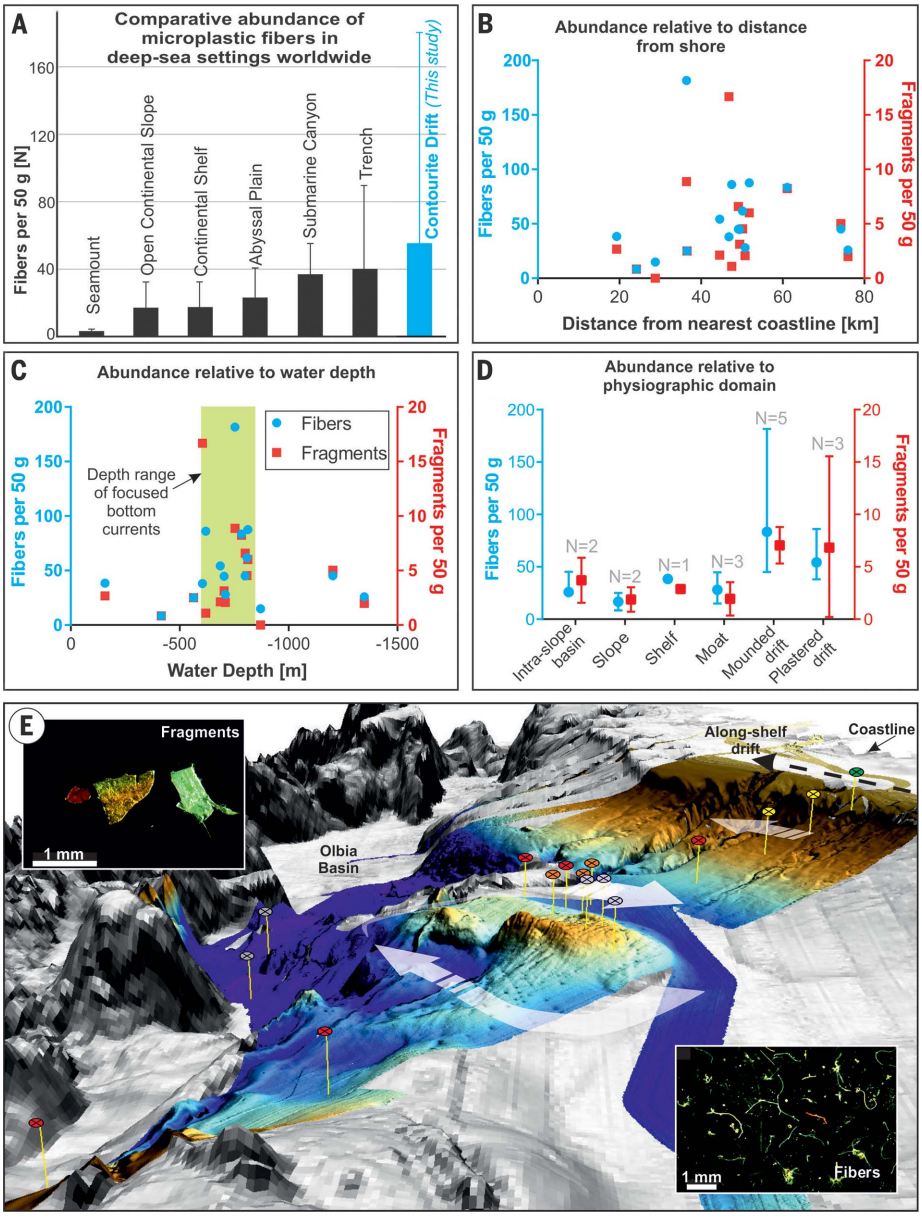
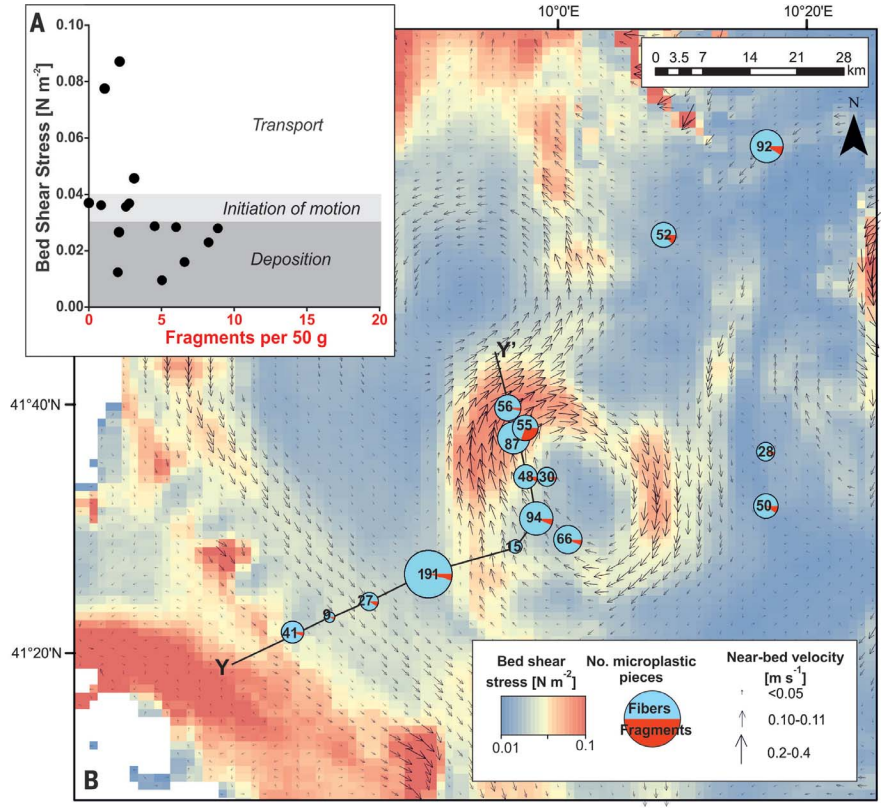


Fig. 2. Global and local abundance of seafloor microplastics. (**A**) Comparison of microplastic fiber abundance in different deep-sea settings worldwide (see supplementary materials), demonstrating the high concentrations observed in the contourite drift deposits in the Tyrrhenian Sea. (**B** to **D**) Results from the Tyrrhenian Sea illustrating that (B) microplastic abundance does not decrease with distance from terrestrial input sources, (C) microplastics appear to be concentrated within a depth range of ~600 to 900 m, and (D) microplastic concentration is biased toward physiographic domains, particularly mounded drifts, and depleted within moats. (**E**) A three-dimensional rendering (perspective view looking toward the southwest) of the regional European Marine Observation and Data Network (EMODNet) bathymetry (grayscale) and autonomous underwater vehicle (AUV) bathymetry (colored) shows the relationship of sediment samples to the local seafloor currents (10× vertical exaggeration). Horizontal distance from bottom left core (red circle) to top left (green circle) is ~100 km. Compare to map view in Fig. 1C. Inset microscope photographs show representative examples of fibers and fragments extracted from the sediment.

Previous Lagrangian modeling of microplastic transport in the Mediterranean [based on a model used to describe global microplastic distributions (*3*)] suggests that waves and sea-surface currents ought to transport microplastics away from the Tyrrhenian Sea (*34*). Therefore, the bottom plastic flux (assuming vertical settling only) in this basin should be one of the lowest: 1.5 to 7 g km^{-2} day^{-1}, compared with a regional maximum of 70 g km^{-2} day^{-1} elsewhere in the Mediterranean (Fig. 1A). If that modeling is correct, microplastic abundances elsewhere in the Mediterranean may be even higher than the values we report here. We suggest, however, that both bottom currents and surface currents are important for the concentration of microplastics, yet near-bed bottom-current circulation is omitted from existing models (*2*, *3*, *8*, *34*). Ocean currents appear to be highly capable of diverting microplastics from shallow to deep water and may be responsible for entraining microplastics transported downslope via submarine channels linked to terrestrial sources (Fig. 5). In enclosed or semienclosed basins, such as the Tyrrhenian Sea and more widely the Mediterranean Sea, circulating contour currents are likely to preferentially accumulate microplastics within contourite drifts. On open continental slopes, contour currents may disperse rather than concentrate microplastics. In such settings, these currents may play a key role in their spatial segregation into hotspots.

We have shown that the overall pattern of bottom currents controls the distribution of microplastics at the seafloor. Numerical modeling and direct measurements in the Tyrrhenian Sea reveal a pronounced seasonal variation in bottom current velocities, with bottom currents being most intense in the winter (*31*). Modeling of microplastic transport shows that some of the near-bed bottom current shear stresses are close to, or in excess of, that required to entrain both fibers and fragments (Figs. 3A and 4). Although low-intensity currents in the summer may allow for the accumulation of microplastics at the seafloor, at some locations (contourite moats, in particular), previously deposited microplastics may be reexhumed as shear stresses exceed the critical boundary for remobilization and suspension (Fig. 3A). More powerful but ephemeral seafloor flows such as gravity currents that have been recorded in deep-sea submarine canyons can reach velocities two orders of magnitude greater than those of the bottom currents in the Tyrrhenian Sea (up to 20 m s^{-1}) and can last for several days (*40*–*42*). These powerful events will undoubtedly flush accumulated microplastics either farther downslope (*43*) or loft them for recirculation by



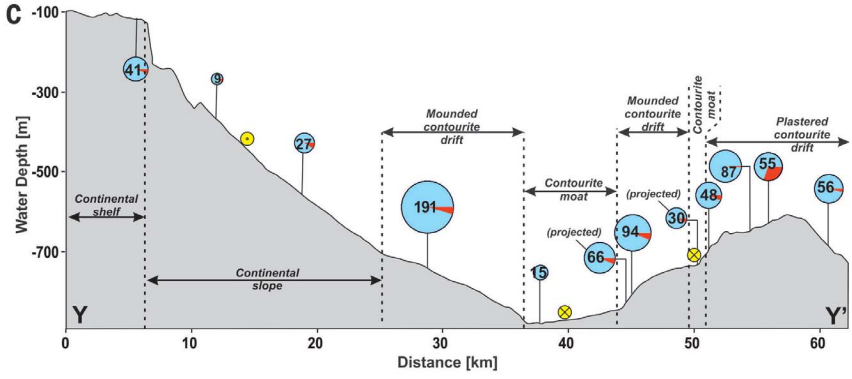


Fig. 3. Influence of bottom currents on the distribution of seafloor microplastics. (**A**) As near-bed shear stresses initially increase, so does the concentration of microplastics; however, when a threshold (~0.04 N m^{-2}) is exceeded, there is a sudden reduction in the abundance of microplastics. This corresponds to the modeled threshold of motion based on empirical approaches (Fig. 4 and fig. S6). (**B**) Hydrodynamic modeling of bottom current circulation shows the formation of seafloor gyres, with corridors of enhanced bed shear stress along the continental slope that are particularly focused within contourite moats and on the flanks of a prominent seamount. Ninetieth percentile for bed shear stress and mean near-bed (bottom current) velocity are shown (see supplementary materials). This focusing of bottom current intensity explains the limited abundance of microplastics on the shelf, upper continental slope, and within contourite moats. (**C**) Zones of lower shear stresses adjacent to these corridors of elevated currents (where mounded contourite drifts form) feature the highest concentrations of microplastics.

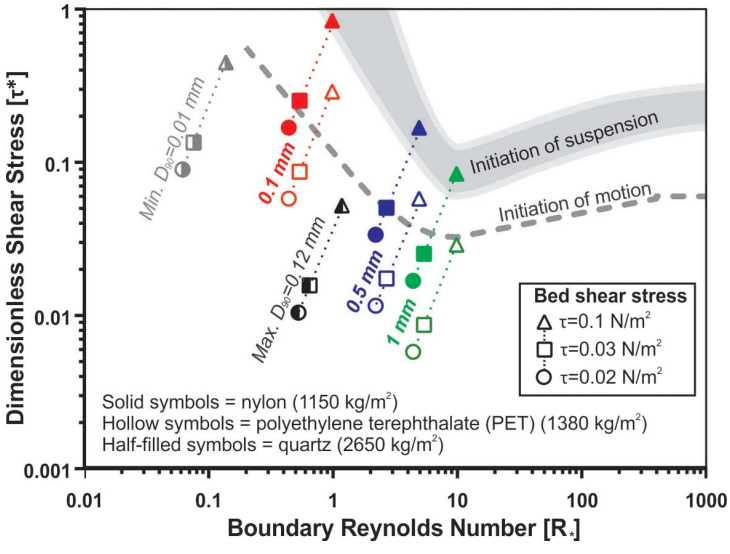


Fig. 4. Relating microplastics to seafloor shear stress. Flow regime diagram to show that the critical shear stress required to move particles [i.e., above the dashed gray line (*51*)] falls between 0.03 and 0.04 N m^{-2}, and that suspension [i.e., within or above the gray shaded area (*52*, *53*)] may occur >0.1 N m^{-2}. Two densities of microplastics are assumed (nylon and polyethylene) on the basis of the results of FTIR analysis. Three particle sizes for microplastics are shown, to represent the general range observed through visual microscopic examination [0.1 mm (red symbols), 0.5 mm (blue), and 1 mm (green) width]. The upper and lower bound measured grain sizes (D_{90}) for the host sediment are also shown with gray symbols.

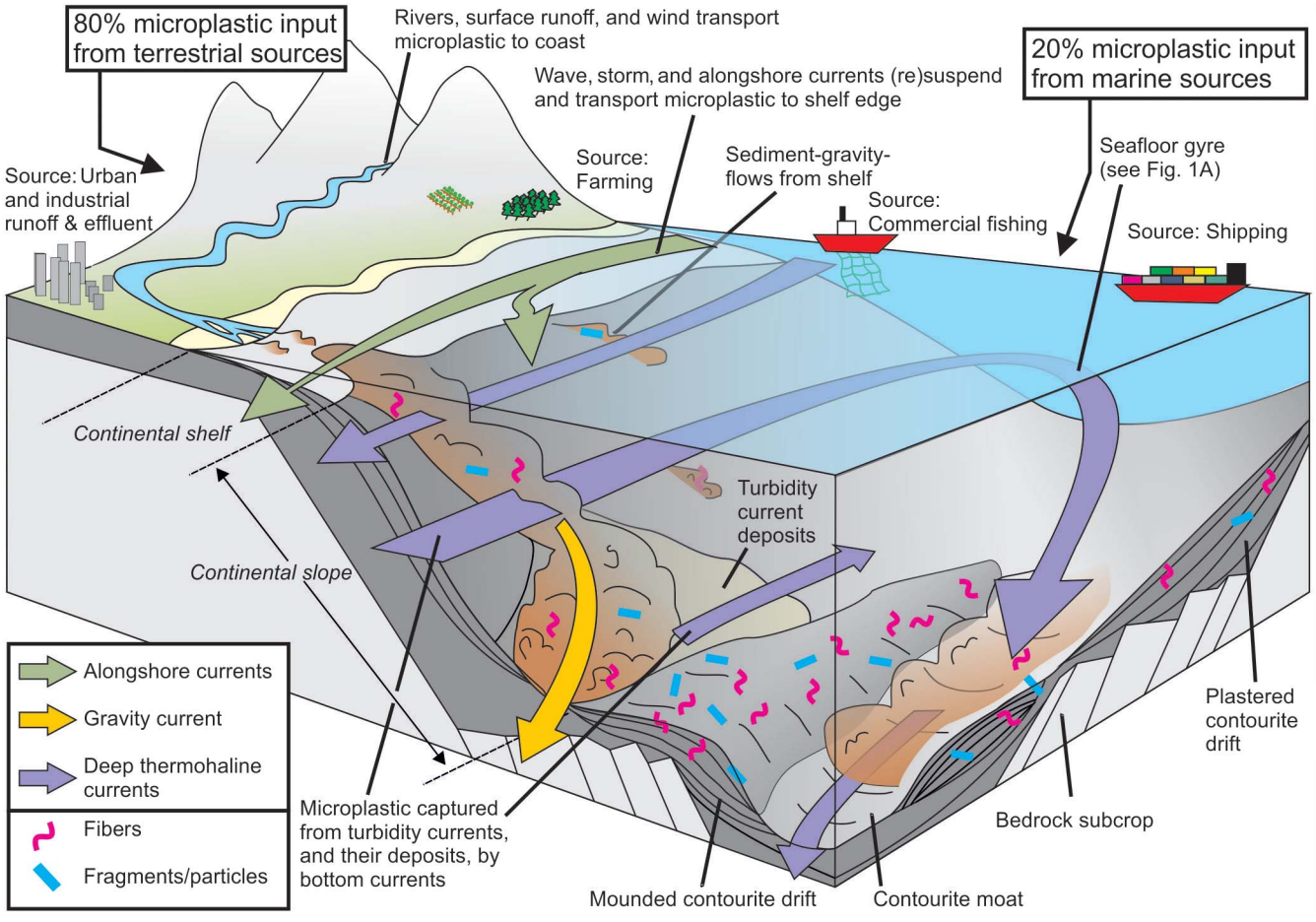


Fig. 5. Bottom currents control the deep-sea fate of microplastics. Schematic diagram illustrating the role of near-bed currents in the transfer, concentration, and storage of microplastics in the deep sea. Along-shelf currents disperse microplastics, powerful gravity flows effectively flush microplastics to the deep sea, while thermohaline-driven bottom currents segregate microplastics into localized hotspots of high concentration. The effectiveness of their long-term sequestration depends on the intensity of subsequent bottom current activity and rate of burial.

thermohaline bottom currents (Fig. 5). Such gravity flows play a globally important role in the lateral transport of lightweight particulate matter such as organic carbon (*41*, *44*); hence, it stands to reason that they should also be important for microplastic transport. Episodic flushing of submarine canyons (*41*, *42*, *45*) suggests that canyons may only be temporary microplastic storage sites (*26*). This is analogous to rivers where flooding can flush high levels of microplastics downstream (*46*).

Bottom currents are efficient conveyors of nutrients and oxygen, and consequently they dictate the location of important biodiversity hotspots (*41*, *47*–*49*). Unfortunately, we show that the same seafloor currents can also transport and emplace microplastics. The highest concentrations of microplastics on the seafloor occur in contourite drifts formed by bottom currents, and their distribution is controlled by spatial variations in current intensity. How effectively microplastics are buried or become reexhumed (and hence become more available for trophic transfer) depends on temporal fluctuations in current intensity. Although there are ongoing efforts to reduce the release of plastics into the environment, our oceans will continue to be affected by the legacy of past waste mismanagement (*4*, *5*, *8*, *50*). Seafloor currents will play a crucial role in the future transfer and storage of microplastics in the deep ocean.

REFERENCES AND NOTES

1. M. L. Taylor, C. Gwinnett, L. F. Robinson, L. C. Woodall, *Sci. Rep.* **6**, 33997 (2016).
2. M. Eriksen *et al.*, *PLOS ONE* **9**, e111913 (2014).
3. E. van Sebille *et al.*, *Environ. Res. Lett.* **10**, 124006 (2015).
4. J. R. Jambeck *et al.*, *Science* **347**, 768–771 (2015).
5. R. Geyer, J. R. Jambeck, K. L. Law, *Sci. Adv.* **3**, e1700782 (2017).
6. L. C. M. Lebreton *et al.*, *Nat. Commun.* **8**, 15611 (2017).
7. R. C. Thompson *et al.*, *Science* **304**, 838–838 (2004).
8. A. A. Koelmans, M. Kooi, K. L. Law, E. van Sebille, *Environ. Res. Lett.* **12**, 114028 (2017).
9. C. A. Choy *et al.*, *Sci. Rep.* **9**, 7843 (2019).
10. L. Van Cauwenberghe, L. Devriese, F. Galgani, J. Robbens, C. R. Janssen, *Mar. Environ. Res.* **111**, 5–17 (2015).
11. A. Vianello *et al.*, *Estuar. Coast. Shelf Sci.* **130**, 54–61 (2013).
12. V. Zitko, M. Hanlon, *Mar. Pollut. Bull.* **22**, 41–42 (1991).
13. S. A. Mason *et al.*, *Environ. Pollut.* **218**, 1045–1054 (2016).
14. M. A. Browne *et al.*, *Environ. Sci. Technol.* **45**, 9175–9179 (2011).
15. A. L. Andrady, *Mar. Pollut. Bull.* **62**, 1596–1605 (2011).
16. X. Tubau *et al.*, *Prog. Oceanogr.* **134**, 379–403 (2015).
17. E. D. Goldberg, *Environ. Technol.* **18**, 195–201 (1997).
18. A. Ballent, A. Purser, P. de Jesus Mendes, S. Pando, L. Thomsen, *Biogeosciences Discuss.* **9**, 18755–18798 (2012).
19. C. K. Pham *et al.*, *PLOS ONE* **9**, e95839 (2014).
20. L. C. Woodall *et al.*, *R. Soc. Open Sci.* **1**, 140317 (2014).
21. V. Fischer, N. O. Elsner, N. Brenke, E. Schwabe, A. Brandt, *Deep Sea Res. Part II Top. Stud. Oceanogr.* **111**, 399–405 (2015).
22. W. Courtene-Jones, B. Quinn, S. F. Gary, A. O. M. Mogg, B. E. Narayanaswamy, *Environ. Pollut.* **231**, 271–280 (2017).
23. A. J. Underwood, M. G. Chapman, M. A. Browne, *Anal. Methods* **9**, 1332–1345 (2017).
24. A. R. Thurber *et al.*, *Biogeosciences* **11**, 3941–3963 (2014).
25. D. K. Barnes, F. Galgani, R. C. Thompson, M. Barlaz, *Philos. Trans. R. Soc. London Ser. B* **364**, 1985–1998 (2009).
26. A. Ballent, S. Pando, A. Purser, M. F. Juliano, L. Thomsen, *Biogeosciences* **10**, 7957–7970 (2013).
27. X. Peng *et al.*, *Geochem. Perspect. Lett.* **9**, 1–5 (2018).
28. I. E. Chubarenko *et al.*, in *Microplastic Contamination in Aquatic Environments* (Elsevier, 2018), pp. 175–223.
29. M. Rebesco, F. J. Hernández-Molina, D. Van Rooij, A. Wåhlin, *Mar. Geol.* **352**, 111–154 (2014).
30. S. Vignudelli, P. Cipollini, M. Astraldi, G. P. Gasparini, G. Manzella, *J. Geophys. Res. Oceans* **105**, 19649–19663 (2000).

31. E. Miramontes *et al*., *Geomorphology* **333**, 43–60 (2019).
32. G. Dalla Valle, F. Gamberi, *Mar. Geol.* **286**, 95–105 (2011).
33. I. N. McCave, *Mar. Geol.* **390**, 89–93 (2017).
34. S. Liubartseva, G. Coppini, R. Lecci, E. Clementi, *Mar. Pollut. Bull.* **129**, 151–162 (2018).
35. C. Cencini, *J. Coast. Res.* **14**, 774–793 (1998).
36. D. A. Kroodsma *et al*., *Science* **359**, 904–908 (2018).
37. M. Bergmann *et al*., *Environ. Sci. Technol.* **51**, 11000–11010 (2017).
38. A. Sanchez-Vidal, R. C. Thompson, M. Canals, W. P. de Haan, *PLOS ONE* **13**, e0207033 (2018).
39. I. A. Kane, M. A. Clare, *Front. Earth Sci.* **7**, 80 (2019).
40. D. J. Piper, P. Cochonat, M. L. Morrison, *Sedimentology* **46**, 79–97 (1999).
41. M. Azpiroz-Zabala *et al*., *Sci. Adv.* **3**, e1700200 (2017).
42. J. J. Mountjoy *et al*., *Sci. Adv.* **4**, eaar3748 (2018).
43. F. Pohl, J. T. Eggenhuisen, I. A. Kane, M. A. Clare, *Environ. Sci. Technol.* **54**, 4180–4189 (2020).
44. V. Galy *et al*., *Nature* **450**, 407–410 (2007).
45. M. Canals *et al*., *Nature* **444**, 354–357 (2006).
46. R. Hurley, J. Woodward, J. J. Rothwell, *Nat. Geosci.* **11**, 251–257 (2018).
47. S. J. Hall, *Environ. Conserv.* **29**, 350–374 (2002).
48. C. Treignier, S. Derenne, A. Saliot, *Org. Geochem.* **37**, 1170–1184 (2006).
49. A. J. Davies *et al*., *Limnol. Oceanogr.* **54**, 620–629 (2009).
50. D. Xanthos, T. R. Walker, *Mar. Pollut. Bull.* **118**, 17–26 (2017).
51. A. Shields, thesis, Technische Universität Berlin (Preussischen Versuchsanstalt für Wasserbau, 1936).
52. L. C. van Rijn, *J. Hydraul. Eng.* **110**, 1613–1641 (1984).
53. Y. Niño, F. Lopez, M. Garcia, *Sedimentology* **50**, 247–263 (2003).
54. I. Kane *et al*., Data from: Seafloor microplastic hotspots controlled by deep-sea circulation, Dryad (2020); https://doi.org/10.5061/dryad.tht76hdwf.

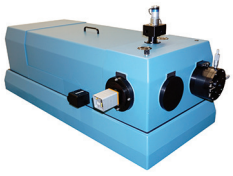
ACKNOWLEDGMENTS

We thank T. Bishop and J. Moore in the Department of Geography at the University of Manchester for help with a range of analyses. We thank the staff at the British Ocean Sediment Core Research Facility (BOSCORF) for access to sediment cores. We thank the GALSI PROJECT for access to survey data. The constructive comments of D. Piper and two anonymous reviewers are gratefully acknowledged. **Funding:** M.A.C. was supported by the CLASS program (NERC grant NE/R015953/1). **Author contributions:** I.A.K. and M.A.C. conceived of and designed the study and subsampled the core samples. I.A.K. carried out the microplastic extraction and analysis. M.A.C. and I.A.K. analyzed seafloor and subsurface data and integrated microplastics concentrations with modeling outputs. E.M. and P.G. modeled the seafloor circulation patterns. F.P. integrated microplastic and sediment transport modeling work. R.W. analyzed FTIR spectra. All authors contributed to writing the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The data that support the findings of this study are available from Dryad (*54*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6495/1140/suppl/DC1
Materials and Methods
Figs. S1 to S5
Table S1
References (*55*–*72*)
Data S1

16 December 2019; accepted 9 April 2020
Published online 30 April 2020
10.1126/science.aba5899



High-Resolution Infrared Spectral Imaging

McPherson dispersive spectrometers measure and tune specific wavelengths of light. They are delivered with diffraction gratings selected for your application. Grating groove density and optimum blaze wavelengths range from ultraviolet and visible wavelengths of 0.2 µm–20 µm long-wave infrared (LWIR). Gold, silver, and other optical coatings enhance performance in the IR. These spectrometers also provide ample, accessible space and a mechanically robust platform for mounting IR arrays, increasing utility for spectroscopy and imaging applications in chemistry, biology, aerospace, and defense.

McPherson
For info: 800-255-1055
www.mcphersoninc.com

Sterilization Systems

AMS Biotechnology offers a trio of sterilization systems suitable for disinfecting everything from small laboratory instruments to large research facilities. These compact sterilizers use a mixture of ozone and UV to kill pathogenic organisms, including viruses, bacteria, and fungi. The CoolCLAVE Plus is a unique, innovative laboratory bench top sterilizer that is highly effective in keeping sterile, or sterilizing contaminated pipettes, pipette tips, gloves, plates, small instruments, and even personal items such as keys, glasses, wallets, and small purses. The Ozilla Ozone Sterilizer is a high-quality, compact, and purpose-built ozone gas generator that can eliminate airborne and surface contaminants and germs. The FirstResponder Sterilizer is a portable system that kills 99.999% of 650 different kinds of pathogenic organisms (bacteria and fungi) in as little as 60 min and can be used to sterilize ambulances, vehicles, and medical research facilities.

AMS Biotechnology
For info: 617-945-5033
www.amsbio.com/sterilization_systems.aspx

Anti-Ipilimumab Antibodies

Bio-Rad Laboratories introduces recombinant, monoclonal anti-idiotypic antibodies, and drug–target complex binders that are specific for ipilimumab (Yervoy) and the ipilimumab/CTLA-4 drug–target complex. These highly specific, high-affinity antibodies are suitable for bioanalytical and patient drug-level monitoring assays for the immune checkpoint inhibitor ipilimumab. Ipilimumab activates the immune system by inhibiting cytotoxic T lymphocyte–associated antigen 4 (CTLA-4), a protein receptor that hinders the destruction of cancerous cells, to treat cancers such as melanoma. The four anti-ipilimumab antibodies inhibit the binding of the drug to its target, CTLA-4, and therefore detect free drug. These antibodies are suitable for developing pharmacokinetic (PK) bridging ELISAs and for use as surrogate positive controls (or reference standards) for anti-drug antibody assays. The drug–target complex binder specifically recognizes the ipilimumab/CTLA-4 complex and can be used as the detection antibody in a PK antigen capture assay as an alternative format to the bridging ELISA.

Bio-Rad Laboratories
For info: 800-424-6723
www.bio-rad-antibodies.com/anti-ipilimumab-antibody-yervoy.html

Animal Venom Toxin Antibodies

Antibodies targeting various snake and spider toxins are now commercially available in newly engineered formats, providing valuable research tools for scientists worldwide striving to produce better antivenom treatments. Originally developed by the laboratory of Andreas Hougaard Laustsen at the Technical University of Denmark (DTU), the antibodies target cytotoxins and phospholipases A2 from African cobras, latrotoxin from widow spiders, and myotoxin II from lancehead vipers. Under the partnership with DTU, Absolute Antibody has engineered the antibodies into new recombinant formats and made them available to a wider scientific audience via an online reagents catalog. For example, they are now available in a human format for use as serological controls, mouse and rabbit formats for detection applications, and sheep formats to enable better comparisons to traditional antivenom. Scientists can also request the antibodies in custom species, isotypes, and subtypes.

Absolute Antibody
For info: 617-377-4057
absoluteantibody.com

Extracellular Matrix for Organoid Culture

Corning Matrigel matrix for organoid culture is an optimized extracellular matrix that reduces the need for time-consuming screening while providing the reproducibility and consistency essential for organoid research. It is verified to support the growth and differentiation of mouse and human organoids, including long-term expansion of mouse small intestinal organoids for more than seven passages with typical organoid budding morphology and marker expression, and growth and differentiation of polarized 3D organoids from primary human airway epithelial cells expressing typical markers. Each lot is measured for matrix stiffness, a property that supports organoid workflow. Additionally, each lot is qualified to form stable "3D dome" structures commonly used in organoid culture. It has also been demonstrated to successfully grow organoids from both healthy and diseased cell origins.

Corning Life Sciences
For info: 800-492-1110
www.corning.com

Intestinal Organoids

DefiniGEN has identified induced pluripotent stem cell (iPSC)-derived intestinal organoids that could be used to help structure in vitro studies of the biology of SARS-CoV-2 infection across cohorts of multiple patients. These organoids display a polarized epithelium and harbor a mixture of cell types normally present in the primary intestinal epithelium barrier in vivo, including goblet cells, Paneth cells, enterocytes, LRG5+ stem cells, and enteroendocrine cells. The organoids polarize, form crypt structures, and grow villi at the apical surface, and are shown to secrete mucus in a similar manner to primary human gut tissue. Several studies have proven that angiotensin-converting enzyme 2 (ACE2) expression in host cells is required for SARS-CoV-2 recognition and infection. Activity of membrane proteases such as transmembrane protease, serine 2 (TMPRSS2) cleaves the coronavirus's spike protein and facilitates the membrane fusion with the host cell. Human intestine is one of the few human tissues with high expression of both ACE2 and TMPRSS2 and is therefore a good candidate to study COVID-19 and the mechanisms of the SARS-CoV-2 infection.

DefiniGEN
For info: 617-674-3260
www.definigen.com/products/intestinal/covid-19

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/about/new-products-section for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

CityU
香港城市大學
City University of Hong Kong

52nd World University

4th World's top 50 Universities under age 50

1st World's Most International Universities

1st Engineering/Technology/ Computer Sciences in Hong Kong

2nd Business School in Asia

Worldwide Search for Talent

City University of Hong Kong is a dynamic, fast-growing university that is pursuing excellence in research and professional education. As a publicly funded institution, the University is committed to nurturing and developing students' talents and creating applicable knowledge to support social and economic advancement. The University has nine Colleges/Schools. As part of its pursuit of excellence, the University aims to recruit **outstanding scholars** from all over the world in various disciplines, including **business, creative media, data science, energy and environment, engineering, humanities and social sciences, law, science, veterinary medicine and life sciences.**

The University welcomes applications and nominations for all faculty positions of Chair Professor, Professor, Associate Professor and Assistant Professor. The remuneration package will be highly competitive, commensurate with qualifications and experience. Interested parties are invited to submit an online application with current curriculum vitae to apply for current openings at http://go.cityu.hk/hrojobus or by email to "hrojob@cityu.edu.hk".

City University of Hong Kong is an equal opportunity employer and we are committed to the principle of diversity. Personal data provided by applicants will be used for recruitment and other employment-related purposes.

Worldwide recognition ranking 52nd, and 4th among top 50 universities under age 50 (QS survey 2020); 1st in the World's Most International Universities (THE survey 2020); 1st in Engineering/Technology/Computer Sciences in Hong Kong (ARWU survey 2016); and 2nd Business School in Asia-Pacific region (UT Dallas survey 2017).

Science Careers online @sciencecareers.org

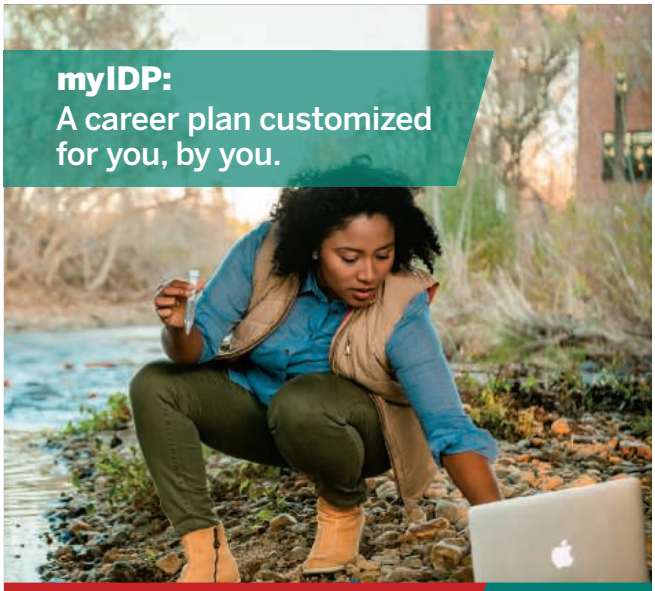
FIND YOUR HAPPIER PLACE.

Find your next job at ScienceCareers.org

There's scientific proof that when you're happy with what you do, you're better at what you do. Access career opportunities, see who's hiring and take advantage of our proprietary career-search tools. Get tailored job alerts, post your resume and manage your applications all in one place: sciencecareers.org

ScienceCareers

FROM THE JOURNAL SCIENCE AAAS



Features in myIDP include:

- Exercises to help you examine your skills, interests, and values.
- A list of 20 scientific career paths with a prediction of which ones best fit your skills and interests.
- A tool for setting strategic goals for the coming year, with optional reminders to keep you on track.
- Articles and resources to guide you through the process.
- Options to save materials online and print them for further review and discussion.
- Ability to select which portion of your IDP you wish to share with advisors, mentors, or others.
- A certificate of completion for users that finish myIDP.

myIDP

Visit the website and start planning today!
myIDP.sciencecareers.org

Science Careers

In partnership with:

UCSF University of California San Francisco

UIC

BURROUGHS WELLCOME FUND

University of Massachusetts Medical School
UMASS

FASEB
Federation of American Societies for Experimental Biology

EPFL

Faculty Position in Soil System Science

at the Ecole polytechnique fédérale de Lausanne (EPFL)

The EPFL School of Architecture, Civil and Environmental Engineering (ENAC) invites applications for a faculty position in its Institute of Environmental Engineering (IIE) at the tenure track level **(Assistant Professor)**.

We encourage applications from experimentalists (field or laboratory) whose vadose zone research interests extend across scales, can inform model development and couple physical, chemical and biological processes. Example research domains include: response of soils to global change, microbial biogeochemistry of soils, role of soils in global greenhouse gas fluxes, soil biodiversity and its interactions with ecosystem functioning, and soil as a provider of ecosystem services. This professorship will interact with other research groups in IIE related to hydrology, ecology, chemistry, microbiology and atmospheric science.

We seek an outstanding individual who will lead an internationally recognized and impactful research program that extends and leverages the opportunities offered by EPFL. The professor will be committed to excellence in research and in undergraduate and graduate level teaching, and will contribute to the teaching program in Environmental Engineering, which views basic and translational research as the foundation for environmental adaption and engineering design.

EPFL is a growing and well-funded institution fostering excellence and diversity. It is well equipped with experimental and computational infrastructure, and offers a fertile environment for research collaboration between different disciplines. The EPFL environment is multilingual and multicultural, with English serving as a common interface. EPFL offers internationally competitive start-up resources, salaries and benefits. Besides its main Lausanne campus, EPFL operates antenna sites across Western Switzerland, in Fribourg, Geneva, Neuchâtel and Sion. This professorship will be located at Sion.

The following documents are requested in PDF format: cover letter including a statement of motivation, curriculum vitae, publication list, concise statements of research and teaching interests (maximum of 5 pages for each statement) as well as the names and addresses, including emails, of at least three references (may be contacted at a later stage). Applications should be uploaded to the EPFL recruitment web site:

https://facultyrecruiting.epfl.ch/position/21823489

Formal evaluation of the applications will begin on **1st July 2020**. The search will continue until the position is filled.

Further enquiries should be made to the Chair of the Search Committee:
Prof. D. Andrew Barry
Director of the Environmental Engineering Institute
E-mail: SearchSoilScience@epfl.ch

For additional information on EPFL, please consult:
www.epfl.ch/en
www.epfl.ch/schools/enac
www.epfl.ch/schools/enac/research/environmental-engineering-institute-iie
www.epfl.ch/schools/enac/education/environmental

EPFL is an equal opportunity employer and family friendly university.
It is committed to increasing the diversity of its faculty.
It strongly encourages women to apply.

By **Guilherme Santos**

No place like home

Our lives changed tracks the day our daughter was born. My wife and I felt settled in the United Kingdom—we'd even bought a house 6 months earlier—but the arrival of our daughter forced a moment of reflection. We wondered whether we should move back to our native Brazil. I remember looking down at our newborn baby and thinking about how different her life would be there. In our native country, we'd be closer to our family and culture. But in the United Kingdom, I had the funding and lab resources I needed for my costly and highly specialized research program. Would such a move sink my career?

We had left Brazil 10 years earlier so that I could spend the last year of my Ph.D. program in France. Meanwhile, my wife started a Ph.D. in the United Kingdom. We maintained a long-distance relationship for the first few years. But by the time our daughter was born, we'd settled down in the same location; I had a research assistant position and my wife had a postdoc.

I was in an amazing scientific environment, working next to Nobel laureates. But I had been away from my extended family and my culture for too long. I missed acai, acarajé, pão de queijo, samba, bossa nova, capoeira, Brazilian jujitsu, and sunny skies. I worried that I might never return to Brazil.

My wife missed our homeland as well. As a mathematician, it was also easier for her to imagine doing her work in Brazil; unlike me, she didn't need fancy lab equipment and expensive reagents. So with a newborn baby at home, we started to ask ourselves whether a move was in order.

Several events over the next few months nudged us further in the same direction. We heard that new professor positions had opened up at a university in our hometown, in departments that matched both of our research programs. Then, a few weeks later, I received a call in the wee hours of the morning from my brother, who told me that our father had passed away. The great happiness I had felt about my daughter's birth suddenly gave way to a deep sadness. I felt that I was losing precious time with my loved ones.

After that, we had no doubt it was the right time to return to Brazil. We applied for the professorships and both received offers. It felt like our personal and professional lives were falling into place.

"I had been away from my extended family and my culture for too long."

It wasn't easy getting started with my research in Brazil, however. I didn't receive any startup funding for my lab, so I couldn't recruit grad students or hire postdocs. I was able to perform a few experiments myself thanks to reagents I had brought with me from my lab in the United Kingdom, as well as equipment I shared with other professors. Eventually, I secured Brazilian funding, which allowed me to form a team of bright undergraduate and graduate students and buy reagents and lab equipment on my own.

Nine years later, I still haven't been able to perform the kind of experiments that U.K. funding would have allowed. And the antiscience views of the current Brazilian government have added to my worries about my squeezed budget.

But looking back, I'm happy that my wife and I put our personal well-being first and made the decision to move back to our home country. We've been able to have big meals with our extended family on weekends. We had a second child, and our kids are growing up close to their cousins who are around the same age. And I was able to resume Brazilian jujitsu training, which I hadn't been able to do during my years abroad.

We are living the lives we wanted to live—working on rewarding research while staying true to what we feel is best for ourselves and for our family. Any scientist from a developing country who is considering a move home has to weigh the pros and cons, and the situation will be different for everyone. But for us, returning home was worth the costs. ■

Guilherme Santos is an associate professor at the University of Brasília. Do you have an interesting career story to share? Send it to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

Published by AAAS